GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 7, 2001, 16:20:07 ; Search time 1335.93 Seconds
(without alignments)
220.788 Million cell updates/sec

Title: US-09-601-561-1
Perfect score: 20
Sequence: 1 ttgstcaaaatttgascaaa 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		18	90.0	37067	93	HSQ9D13	AL109764 Homo sapi
2		18	90.0	78447	92	HSF1614	AL109761 Homo sapi
3	c	18	90.0	143893	76	AC079139	AC079139 Homo sapi
4		18	90.0	170800	69	AC024411	AC024411 Homo sapi
5		18	90.0	180017	75	AC073277	AC073277 Homo sapi
6	c	18	90.0	340000	90	AP001671	AP001671 Homo sapi
7		17.6	88.0	8000	64	AC015191	AC015191 Drosophil
8	c	17.6	88.0	14922	6	CELY55F3AR	AC024827 Caenorhab
9	c	17.6	88.0	22964	6	CELY19D10B	AC006723 Caenorhab
10		17.6	88.0	35500	5	CEC54C6	Z77131 Caenorhabdi
11		17.6	88.0	42728	63	AC012726	AC012726 Drosophil

	12	17.6	88.0	71585	61	AC009745 AC009745 Drosophil
c	13	17.6	88.0	92237	60	AC008307 AC008307 Drosophil
	14	17.6	88.0	95668	61	AC009254 AC009254 Drosophil
	15	17.6	88.0	108399	80	AL356736 AL356736 Homo sapi
	16	17.6	88.0	110000	83	CEY39B6_1 Continuation (2 of
c	17	17.6	88.0	143475	83	HSDJ543C6 AL109926 Homo sapi
c	18	17.6	88.0	159425	79	AL158081 AL158081 Homo sapi
c	19	17.6	88.0	161641	69	AC024697 AC024697 Homo sapi
	20	17.6	88.0	163198	79	AL161634 AL161634 Homo sapi
c	21	17.6	88.0	174662	6	CEY59A8B AL132898 Caenorhab
c	22	17.6	88.0	176950	80	AL357273 AL357273 Homo sapi
	23	17.6	88.0	186747	65	AC018605 AC018605 Homo sapi
c	24	17.6	88.0	186762	72	AC040168 AC040168 Homo sapi
c	25	17.6	88.0	203987	76	AC079409 AC079409 Homo sapi
	26	17.6	88.0	207022	6	CEY39B6B AL132896 Caenorhab
c	27	17.6	88.0	233034	60	AC007495 AC007495 Homo sapi
c	28	17.6	88.0	293024	83	CEY59A8 Z98870 Caenorhabdi
c	29	17.6	88.0	298705	4	AE003510 AE003510 Drosophil
	30	17.6	88.0	299300	60	AC006881 AC006881 Caenorhab
c	31	17.6	88.0	299864	60	AC006702 AC006702 Caenorhab
	32	17.6	88.0	311800	4	AE003781 AE003781 Drosophil
	33	17.6	88.0	338007	60	AC008318 AC008318 Drosophil
c	34	17.2	86.0	171	5	AF215121 AF215121 Conus are
c	35	17.2	86.0	22382	6	CELY46B2A AC006770 Caenorhab
c	36	17	85.0	5552	12	AF133118 AF133118 Oryza sat
c	37	17	85.0	37569	6	CET26H2 Z82055 Caenorhabdi
	38	17	85.0	43255	83	CEC37B6 Z81480 Caenorhabdi
	39	17	85.0	56870	83	CEY39B6_3 Continuation (4 of
	40	17	85.0	106842	83	CEY116F11_8 Continuation (9 of
	41	17	85.0	152848	81	AL450423 AL450423 Homo sapi
c	42	17	85.0	176056	89	AL136359 AL136359 Human DNA
	43	17	85.0	179756	75	AC074039 AC074039 Mus muscu
c	44	17	85.0	183338	77	AC083775 AC083775 Homo sapi
	45	17	85.0	193444	13	AF229199 AF229199 Oryza sat
	46	16.8	84.0	1048	54	CNS07C9V AL438905 T7 end of
	47	16.6	83.0	925	72	AC042158 AC042158 Giardia i
c	48	16.6	83.0	959	71	AC030109 AC030109 Giardia i
	49	16.6	83.0	963	72	AC042455 AC042455 Giardia i
	50	16.6	83.0	1628	14	MCU79767 U79767 Mesembryant

ALIGNMENTS

RESULT 1
HSQ9D13
LOCUS HSQ9D13 37067 bp DNA PRI 14-DEC-1999
DEFINITION Homo sapiens chromosome 21 sequence from Cosmid LLNLc116 9D13 map 21q21 region D21S189-LA297, complete sequence.
ACCESSION AL109764
VERSION AL109764.2 GI:5931932
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 37067)
AUTHORS Ramser,J., Borzym,K., Lehrack,S., Klein,M., Arndt,S., Hildmann,T., Gardiner,K., Yaspo,M.L., Reinhardt,R. and Lehrach,H.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 37067)
AUTHORS MPIMG.
TITLE Direct Submission
JOURNAL Submitted (06-AUG-1999) MPIMG, Abt.Lehrach, Max Planck Institut Fuer Molekulare Genetik, Ihnestrasse 73, Berlin, 14195 Germany
COMMENT On Sep 28, 1999 this sequence version replaced gi:5725299.
contig 01 1. .37067.
FEATURES Location/Qualifiers
source 1. .37067
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="Cosmid LLNLc116 9D13"
/clone_lib="LL21NC02, chromosome 21 specific cosmid library, creator:Pieter de Jong"
/map="21q21"
/note="region between markers D21S189-LA297"
repeat_region complement(2349. .2502)
/note="MER4A"
/rpt_family="LTR/MER4-group"
repeat_region complement(2708. .2826)
/note="MER4C"
/rpt_family="LTR/MER4-group"
repeat_region 2864. .3387
/note="L1MC4"
/rpt_family="LINE/L1"
repeat_region 3394. .3532
/note="L1MC4"
/rpt_family="LINE/L1"
repeat_region 3606. .3907
/note="AluSq"
/rpt_family="SINE/Alu"
repeat_region 3962. .4008
/note="AT_rich"
/rpt_family="Low_complexity"
repeat_region complement(4355. .4545)
/note="MIR"
/rpt_family="SINE/MIR"
repeat_region 5370. .5409
/note="(CATA)n"
/rpt_family="Simple_repeat"
repeat_region complement(5519. .5549)
/note="AT_rich"
/rpt_family="Low_complexity"
repeat_region complement(5564. .5755)
/note="MIR"
/rpt_family="SINE/MIR"
repeat_region 6192. .12630
/note="L1PA2"
/rpt_family="LINE/L1"
repeat_region 14989. .15334
/note="MLT1A1"
/rpt_family="LTR/MaLR"
repeat_region complement(15564. .15602)
/note="U2"
/rpt_family="snRNA"
repeat_region 15896. .15975
/note="AT_rich"
/rpt_family="Low_complexity"
repeat_region complement(15984. .16285)
/note="AluY"
/rpt_family="SINE/Alu"
repeat_region complement(16570. .16619)
/note="MER5A"
/rpt_family="DNA/MER1_type"
repeat_region complement(16829. .16928)
/note="MER45"
/rpt_family="DNA/MER1_type"
repeat_region complement(16939. .17054)
/note="(TAAA)n"
/rpt_family="Simple_repeat"
repeat_region 17502. .17668
/note="MER5A"
/rpt_family="DNA/MER1_type"
repeat_region complement(17824. .17862)
/note="(CA)n"
/rpt_family="Simple_repeat"
repeat_region 19589. .19739
/note="AluJo/FRAM"
/rpt_family="SINE/Alu"
repeat_region complement(20209. .20554)
/note="MLT1A1"
/rpt_family="LTR/MaLR"
repeat_region complement(21983. .22680)
/note="MER67C"
/rpt_family="Other"

```
     repeat_region    22741. .22784
                      /note="Alu"
                      /rpt_family="SINE/Alu"
     repeat_region    22786. .22847
                      /note="L1MA9"
                      /rpt_family="LINE/L1"
     repeat_region    23016. .23259
                      /note="MER39"
                      /rpt_family="Other/MER21_gro"
     repeat_region    complement(23277. .23670)
                      /note="MSTB"
                      /rpt_family="LTR/MaLR"
     repeat_region    23686. .23787
                      /note="MER39"
                      /rpt_family="Other/MER21_gro"
     repeat_region    23787. .24048
                      /note="MER39b_j"
                      /rpt_family="Other/MER21_gro"
     repeat_region    24167. .24262
                      /note="AluSp"
                      /rpt_family="SINE/Alu"
     repeat_region    24280. .25648
                      /note="pTR5"
                      /rpt_family="LTR/Retroviral"
     repeat_region    25671. .25754
                      /note="AluSg/x"
                      /rpt_family="SINE/Alu"
     repeat_region    26075. .26377
                      /note="AluJb"
                      /rpt_family="SINE/Alu"
     repeat_region    26386. .26670
                      /note="AluJo"
                      /rpt_family="SINE/Alu"
     repeat_region    26721. .26783
                      /note="L1"
                      /rpt_family="LINE/L1"
     repeat_region    26868. .27163
                      /note="AluJb"
                      /rpt_family="SINE/Alu"
     repeat_region    27181. .27210
                      /note="AT_rich"
                      /rpt_family="Low_complexity"
     repeat_region    27477. .27765
                      /note="L1P5"
                      /rpt_family="LINE/L1"
     repeat_region    28126. .28193
                      /note="(CATA)n"
                      /rpt_family="Simple_repeat"
     repeat_region    28196. .28560
                      /note="MER39"
                      /rpt_family="Other/MER21_gro"
     repeat_region    28558. .28741
                      /note="MER39b_j"
                      /rpt_family="Other/MER21_gro"
     repeat_region    28822. .28862
                      /note="(CATA)n"
                      /rpt_family="Simple_repeat"
     repeat_region    complement(28998. .29032)
                      /note="AT_rich"
                      /rpt_family="Low_complexity"
     repeat_region    30128. .34316
                      /note="L1PA2"
                      /rpt_family="LINE/L1"
     repeat_region    34319. .34374
                      /note="(GAAAA)n"
                      /rpt_family="Simple_repeat"
     repeat_region    complement(35013. .35117)
                      /note="MER34"
                      /rpt_family="Other/MER21_gro"
     repeat_region    35221. .35275
                      /note="MER4D"
                      /rpt_family="LTR/MER4-group"
     repeat_region    complement(35440. .35535)
                      /note="MER34"
                      /rpt_family="Other/MER21_gro"
     repeat_region    complement(36094. .36448)
                      /note="MER39"
                      /rpt_family="Other/MER21_gro"
     repeat_region    complement(36449. .36613)
                      /note="MER39b_j"
                      /rpt_family="Other/MER21_gro"
     repeat_region    36614. .36915
                      /note="AluY"
                      /rpt_family="SINE/Alu"
     repeat_region    complement(36916. .37067)
                      /note="MER39b_j"
                      /rpt_family="Other/MER21_gro"
BASE COUNT    12597 a    7547 c    7562 g    9361 t
ORIGIN

  Query Match             90.0%;  Score 18;  DB 93;  Length 37067;
  Best Local Similarity   90.0%;  Pred. No. 1.6e+02;
  Matches   18;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ttgstcaaaatttgascaaa 20
           ||| |||||||||||:||||
Db   21375 TTGATCAAAATTTGACCAAA 21394


RESULT   2
HSF1614
LOCUS       HSF1614     78447 bp    DNA             PRI       18-FEB-2000
DEFINITION  Homo sapiens chromosome 21 sequence from PAC RPCI-1 14F16 map 21q21
            region D21S189-LA297, complete sequence.
ACCESSION   AL109761
VERSION     AL109761.3  GI:7018381
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 78447)
  AUTHORS   Ramser,J., Borzym,K., Langer,I., Lehrack,S., Klein,M., Arndt,S.,
            Marquardt,I., Noitz,M., Riesselmann,L., Yaspo,M.L., Reinhardt,R.
            and Lehrach,H.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 78447)
  AUTHORS   MPIMG.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-1999) MPIMG, Abt.Lehrach, Max Planck Institut
            Fuer Molekulare Genetik, Ihnestrasse 73, Berlin, 14195 Germany
COMMENT     On Feb 21, 2000 this sequence version replaced gi:6982091.
            contig 01         1. .78447
            Clone received from Resource Centre of the Human Genome Project at
            the Max-Planck-Institut for Molecular Genetics.
FEATURES             Location/Qualifiers
     source          1. .78447
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /clone="PAC RPCI-1 14F16"
                     /clone_lib="RPCI1,3-5 Human PAC library, originating
                     institute: Roswell Park Cancer Institute, creator: Pieter
                     de Jong, P.Ioannou"
                     /map="21q21"
                     /note="region between markers D21S189-LA297"
BASE COUNT    25001 a  15295 c  15653 g  22498 t
ORIGIN

  Query Match             90.0%;  Score 18;  DB 92;  Length 78447;
  Best Local Similarity   90.0%;  Pred. No. 1.6e+02;
  Matches   18;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ttgstcaaaatttgascaaa 20
```

```
          ||| ||||||||||||:||||
Db  61488 TTGATCAAAATTTGACCAAA 61507


RESULT   3
AC079139/c
LOCUS       AC079139   143893 bp    DNA             HTG       02-SEP-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-477L2, WORKING DRAFT SEQUENCE,
            6 unordered pieces.
ACCESSION   AC079139
VERSION     AC079139.3  GI:9965016
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 143893)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 143893)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-AUG-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 2, 2000 this sequence version replaced gi:9858997.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0477L02
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 139774 bases at least Q40
            Consensus quality: 140896 bases at least Q30
            Consensus quality: 141562 bases at least Q20
            Insert size: 144000; agarose-fp
            Insert size: 142600; sum-of-contigs
            Quality coverage: 7.36 in Q20 bases; agarose-fp
            Quality coverage: 6.84 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1212: contig of 1212 bp in length
            *     1213     1312: gap of unknown length
            *     1313     3091: contig of 1779 bp in length
            *     3092     3191: gap of unknown length
            *     3192     5227: contig of 2036 bp in length
            *     5228     5327: gap of unknown length
            *     5328    25823: contig of 20496 bp in length
            *    25824    25923: gap of unknown length
            *    25924    26716: contig of 793 bp in length
            *    26717    26816: gap of unknown length
            *    26817   143893: contig of 117077 bp in length.
FEATURES             Location/Qualifiers
     source          1. .143893
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-477L2"
     misc_feature    1. .1212
                     /note="assembly_name:Contig16"
     misc_feature    1313. .3091
                     /note="assembly_name:Contig18"
     misc_feature    3192. .5227
                     /note="assembly_name:Contig19"
     misc_feature    5328. .25823
                     /note="assembly_name:Contig20
                     clone_end:SP6
                     vector_side:left"
     misc_feature    25924. .26716
                     /note="assembly_name:Contig11"
     misc_feature    26817. .143893
                     /note="assembly_name:Contig21"
BASE COUNT    44071 a  26366 c  27342 g  45587 t    527 others
ORIGIN


  Query Match             90.0%;  Score 18;  DB 76;  Length 143893;
  Best Local Similarity   90.0%;  Pred. No. 1.6e+02;
  Matches   18;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ttgstcaaaatttgascaaa 20
           ||| ||||||||||:||||
Db   39208 TTGTTCAAAATTTGACCAAA 39189


RESULT   4
AC024411
LOCUS       AC024411   170800 bp    DNA             HTG       25-MAY-2000
DEFINITION  Homo sapiens chromosome 21 clone RP11-383H8 map 21, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC024411
VERSION     AC024411.3  GI:8076906
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170800)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 21, clone RP11-383H8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170800)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7272333.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
```

```
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
          ---------- Project Information
            Center project name: L6873
            Center clone name: 383_H_8
          ---------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 161079 bases at least Q40
            Consensus quality: 165699 bases at least Q30
            Consensus quality: 167817 bases at least Q20
            Insert size: 170000; agarose-fp
            Insert size: 169500; sum-of-contigs
            Quality coverage: 4.1 in Q20 bases; agarose-fp
            Quality coverage: 4.2 in Q20 bases; sum-of-contigs
          ----------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 14 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     1034: contig of 1034 bp in length
          *     1035  1134: gap of      100 bp
          *     1135     2818: contig of 1684 bp in length
          *     2819  2918: gap of      100 bp
          *     2919     5025: contig of 2107 bp in length
          *     5026  5125: gap of      100 bp
          *     5126     6457: contig of 1332 bp in length
          *     6458  6557: gap of      100 bp
          *     6558     9772: contig of 3215 bp in length
          *     9773  9872: gap of      100 bp
          *     9873    14119: contig of 4247 bp in length
          *    14120 14219: gap of      100 bp
          *    14220    18249: contig of 4030 bp in length
          *    18250 18349: gap of      100 bp
          *    18350    23124: contig of 4775 bp in length
          *    23125 23224: gap of      100 bp
          *    23225    26842: contig of 3618 bp in length
          *    26843 26942: gap of      100 bp
          *    26943    43018: contig of 16076 bp in length
          *    43019 43118: gap of      100 bp
          *    43119    63582: contig of 20464 bp in length
          *    63583 63682: gap of      100 bp
          *    63683    88966: contig of 25284 bp in length
          *    88967 89066: gap of      100 bp
          *    89067   126431: contig of 37365 bp in length
          *   126432 126531: gap of      100 bp
          *   126532   170800: contig of 44269 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170800
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21"
                     /clone="RP11-383H8"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1034
                     /note="assembly_fragment"
     misc_feature    1135. .2818
                     /note="assembly_fragment"
     misc_feature    2919. .5025
                     /note="assembly_fragment"
     misc_feature    5126. .6457
                     /note="assembly_fragment"
     misc_feature    6558. .9772
                     /note="assembly_fragment"
     misc_feature    9873. .14119
                     /note="assembly_fragment"
     misc_feature    14220. .18249
                     /note="assembly_fragment"
     misc_feature    18350. .23124
                     /note="assembly_fragment"
     misc_feature    23225. .26842
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    26943. .43018
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    43119. .63582
                     /note="assembly_fragment"
     misc_feature    63683. .88966
                     /note="assembly_fragment"
     misc_feature    89067. .126431
                     /note="assembly_fragment"
     misc_feature    126532. .170800
                     /note="assembly_fragment"
BASE COUNT    53037 a  33679 c  33422 g  49351 t   1311 others
ORIGIN


  Query Match             90.0%;  Score 18;  DB 69;  Length 170800;
  Best Local Similarity   90.0%;  Pred. No. 1.6e+02;
  Matches   18;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          ||| |||||||||||:||||
Db 157245 TTGATCAAAATTTGACCAAA 157264


RESULT   5
AC073277
LOCUS       AC073277   180017 bp    DNA             HTG       18-JUN-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-657G5, WORKING DRAFT SEQUENCE,
            28 unordered pieces.
ACCESSION   AC073277
VERSION     AC073277.2  GI:8571830
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180017)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180017)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 18, 2000 this sequence version replaced gi:8469052.

            ---------- Genome Center ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 28 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1246: contig of 1246 bp in length
            *     1247     1346: gap of unknown length
            *     1347     3076: contig of 1730 bp in length
            *     3077     3176: gap of unknown length
            *     3177     4590: contig of 1414 bp in length
            *     4591     4690: gap of unknown length
```

```
            *     4691     6416: contig of 1726 bp in length
            *     6417     6516: gap of unknown length
            *     6517     7902: contig of 1386 bp in length
            *     7903     8002: gap of unknown length
            *     8003     9781: contig of 1779 bp in length
            *     9782     9881: gap of unknown length
            *     9882    12537: contig of 2656 bp in length
            *    12538    12637: gap of unknown length
            *    12638    15059: contig of 2422 bp in length
            *    15060    15159: gap of unknown length
            *    15160    17468: contig of 2309 bp in length
            *    17469    17568: gap of unknown length
            *    17569    20886: contig of 3318 bp in length
            *    20887    20986: gap of unknown length
            *    20987    24569: contig of 3583 bp in length
            *    24570    24669: gap of unknown length
            *    24670    27911: contig of 3242 bp in length
            *    27912    28011: gap of unknown length
            *    28012    31184: contig of 3173 bp in length
            *    31185    31284: gap of unknown length
            *    31285    35127: contig of 3843 bp in length
            *    35128    35227: gap of unknown length
            *    35228    38363: contig of 3136 bp in length
            *    38364    38463: gap of unknown length
            *    38464    41653: contig of 3190 bp in length
            *    41654    41753: gap of unknown length
            *    41754    46937: contig of 5184 bp in length
            *    46938    47037: gap of unknown length
            *    47038    52755: contig of 5718 bp in length
            *    52756    52855: gap of unknown length
            *    52856    57749: contig of 4894 bp in length
            *    57750    57849: gap of unknown length
            *    57850    64952: contig of 7103 bp in length
            *    64953    65052: gap of unknown length
            *    65053    72484: contig of 7432 bp in length
            *    72485    72584: gap of unknown length
            *    72585    82804: contig of 10220 bp in length
            *    82805    82904: gap of unknown length
            *    82905    94700: contig of 11796 bp in length
            *    94701    94800: gap of unknown length
            *    94801   107766: contig of 12966 bp in length
            *   107767   107866: gap of unknown length
            *   107867   121590: contig of 13724 bp in length
            *   121591   121690: gap of unknown length
            *   121691   136975: contig of 15285 bp in length
            *   136976   137075: gap of unknown length
            *   137076   160258: contig of 23183 bp in length
            *   160259   160358: gap of unknown length
            *   160359   180017: contig of 19659 bp in length.
FEATURES             Location/Qualifiers
     source          1. .180017
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-657G5"
     misc_feature    1. .1246
                     /note="assembly_name:Contig7"
     misc_feature    1347. .3076
                     /note="assembly_name:Contig8"
     misc_feature    3177. .4590
                     /note="assembly_name:Contig9"
     misc_feature    4691. .6416
                     /note="assembly_name:Contig10"
     misc_feature    6517. .7902
                     /note="assembly_name:Contig11"
     misc_feature    8003. .9781
                     /note="assembly_name:Contig12"
     misc_feature    9882. .12537
                     /note="assembly_name:Contig13"
     misc_feature    12638. .15059
                     /note="assembly_name:Contig14"
     misc_feature    15160. .17468
                     /note="assembly_name:Contig16"
     misc_feature    17569. .20886
                     /note="assembly_name:Contig17"
     misc_feature    20987. .24569
                     /note="assembly_name:Contig18"
     misc_feature    24670. .27911
                     /note="assembly_name:Contig19"
     misc_feature    28012. .31184
                     /note="assembly_name:Contig20"
     misc_feature    31285. .35127
                     /note="assembly_name:Contig21"
     misc_feature    35228. .38363
                     /note="assembly_name:Contig22"
     misc_feature    38464. .41653
                     /note="assembly_name:Contig23
                     clone_end:SP6
                     vector_side:right"
     misc_feature    41754. .46937
                     /note="assembly_name:Contig24"
     misc_feature    47038. .52755
                     /note="assembly_name:Contig25"
     misc_feature    52856. .57749
                     /note="assembly_name:Contig26"
     misc_feature    57850. .64952
                     /note="assembly_name:Contig27"
     misc_feature    65053. .72484
                     /note="assembly_name:Contig28"
     misc_feature    72585. .82804
                     /note="assembly_name:Contig29"
     misc_feature    82905. .94700
                     /note="assembly_name:Contig30"
     misc_feature    94801. .107766
                     /note="assembly_name:Contig31"
     misc_feature    107867. .121590
                     /note="assembly_name:Contig32"
     misc_feature    121691. .136975
                     /note="assembly_name:Contig33"
     misc_feature    137076. .160258
                     /note="assembly_name:Contig34"
     misc_feature    160359. .180017
                     /note="assembly_name:Contig35"
BASE COUNT    56437 a  33832 c  33071 g  53973 t   2704 others
ORIGIN


  Query Match             90.0%;  Score 18;  DB 75;  Length 180017;
  Best Local Similarity   90.0%;  Pred. No. 1.6e+02;
  Matches   18;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ttgstcaaaatttgascaaa 20
           ||| |||||||||||:||||
Db   29572 TTGTTCAAAATTTGACCAAA 29591


RESULT   6
AP001671/c
LOCUS       AP001671   340000 bp    DNA             PRI       30-MAY-2000
DEFINITION  Homo sapiens genomic DNA, chromosome 21q, section 15/105.
ACCESSION   AP001671 AL163216 BA000005
VERSION     AP001671.1  GI:7768665
KEYWORDS    .
SOURCE      Homo sapiens DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M. and Schudy,A.
  TITLE     The DNA sequence of human chromosome 21. The chromosome 21 mapping
            and sequencing consortium
  JOURNAL   Nature 405 (6784), 311-319 (2000)
```

```
  MEDLINE   20289799
REFERENCE   2  (bases 1 to 340000)
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Hennig,S., Riesselmann,L., Dagand,E., Wehrmaeyer,S., Borzym,K.,
            Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
    TITLE   Direct Submission
  JOURNAL   Submitted (10-APR-2000) to the DDBJ/EMBL/GenBank databases. The
            Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic
            Sciences Center, Human Genome Research Group * Institute of
            Molecular Biotechnology, Genome Analysis * Keio University School
            of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome
            Analysis * Max-Planck Institute for Molecular Genetics (addresses
            see below)
  COMMENT   On May 30, 2000 this sequence version replaced gi:7717265.
            The chromosome 21 mapping and sequencing consortium consisting of
            * RIKEN Genomic Sciences Center, Human Genome Research Group, *
            Sagamihara 228-8555, Japan,
            * e.mail: hattori@gsc.riken.go.jp
            * URL: http://hgp.gsc.riken.go.jp/
            and
            * Institute of Molecular Biotechnology, Genome Analysis, *
            Beutenbergstrasse 11, D-07745 Jena, Germany,
            * e.mail: gscj-submit@genome.imb-jena.de
            * URL: http://genome.imb-jena.de/
            and
            * Keio University School of Medicine, Molecular Biology, * Tokyo
            160-8582, Japan,
            * e.mail: nshimizu@dmb-med.keio.ac.jp
            * URL: http://www.dmb.med.keio.ac.jp/
            and
            * GBF, Dept. of Genome Analysis,
            * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
            info.genome@gbf.de
            * URL: http://genome.gbf.de/
            and
            * Max-Planck Institute for Molecular Genetics,
            * Ihnestrasse 73, D-14195 Berlin, Germany,
            * e.mail: info-chr21@molgen.mpg.de
            * URL: http://chr21.rz-berlin.mpg.de/
            AL163216: Submitted (10-Apr-2000).
FEATURES             Location/Qualifiers
     source          1. .340000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
     source          <1. .36248
                     /note="Accession No. AP000432"
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21.1"
                     /clone="R396A17, 5' partial"
                     /clone_lib="RPCI-11 BAC library"
     repeat_region   complement(3292. .3734)
                     /note="L1MA9"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   complement(3755. .4445)
                     /note="THE1B-internal"
                     /rpt_family="LTR/MaLR"
                     /rpt_type=DISPERSED
     repeat_region   complement(4446. .4744)
                     /note="THE1B"
                     /rpt_family="LTR/MaLR"
                     /rpt_type=DISPERSED
     repeat_region   complement(6107. .6389)
                     /note="AluSx"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   complement(6590. .6926)
                     /note="MLT1C"
                     /rpt_family="LTR/MaLR"
                     /rpt_type=DISPERSED
     repeat_region   6929. .7566
                     /note="L1PB3"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   complement(7610. .7829)
                     /note="AluSx"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   7834. .8505
                     /note="L1PB3"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   complement(8506. .8544)
                     /note="THE1C"
                     /rpt_family="LTR/MaLR"
                     /rpt_type=DISPERSED
     repeat_region   8545. .8840
                     /note="AluY"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   complement(8841. .9188)
                     /note="THE1C"
                     /rpt_family="LTR/MaLR"
                     /rpt_type=DISPERSED
     repeat_region   9189. .9350
                     /note="L1PB3"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   9359. .10350
                     /note="SVA"
                     /rpt_family="LTR/Retroviral"
                     /rpt_type=DISPERSED
     repeat_region   10351. .10596
                     /note="SVA"
                     /rpt_family="LTR/Retroviral"
                     /rpt_type=DISPERSED
     repeat_region   10597. .11085
                     /note="SVA"
                     /rpt_family="LTR/Retroviral"
                     /rpt_type=DISPERSED
     repeat_region   11150. .11325
                     /note="SVA"
                     /rpt_family="LTR/Retroviral"
                     /rpt_type=DISPERSED
     repeat_region   11326. .11828
                     /note="SVA"
                     /rpt_family="LTR/Retroviral"
                     /rpt_type=DISPERSED
     repeat_region   11891. .13355
                     /note="L1PB3"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   13356. .13656
                     /note="AluSg"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   13657. .13895
                     /note="L1PB3"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   complement(13900. .14042)
                     /note="MLT1C"
                     /rpt_family="LTR/MaLR"
```

```
                     /rpt_type=DISPERSED
     repeat_region   14252. .14548
                     /note="L2"
                     /rpt_family="LINE/L2"
                     /rpt_type=DISPERSED
     repeat_region   14549. .14689
                     /note="FLAM_A"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   14690. .14822
                     /note="L2"
                     /rpt_family="LINE/L2"
                     /rpt_type=DISPERSED
     repeat_region   complement(14816. .14949)
                     /note="MER5A"
                     /rpt_family="DNA/MER1_type"
                     /rpt_type=DISPERSED
     repeat_region   15003. .15348
                     /note="L2"
                     /rpt_family="LINE/L2"
                     /rpt_type=DISPERSED
     repeat_region   15350. .15406
                     /note="MER50"
                     /rpt_family="LTR/MER4-group"
                     /rpt_type=DISPERSED
     repeat_region   complement(15425. .15578)
                     /note="AluY"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   15579. .15729
                     /note="MER50"
                     /rpt_family="LTR/MER4-group"
                     /rpt_type=DISPERSED
     repeat_region   15793. .15833
                     /note="(TAAA)n"
                     /rpt_family="Simple_repeat"

  Query Match             90.0%;  Score 18;  DB 90;  Length 340000;
  Best Local Similarity   90.0%;  Pred. No. 1.5e+02;
  Matches   18;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          ||| |||||||||||:||||
Db  51764 TTGATCAAAATTTGACCAAA 51745


RESULT   7
AC015191
LOCUS       AC015191     8000 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC015191
VERSION     AC015191.1  GI:6436144
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 8000)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10213666 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .8000
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     2586 a   1471 c   1442 g   2501 t
ORIGIN


  Query Match             88.0%;  Score 17.6;  DB 64;  Length 8000;
  Best Local Similarity   85.0%;  Pred. No. 2.7e+02;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||:|||||||||||: |||
Db   7598 TTGCTCAAAATTTGAGAAAA 7617


RESULT   8
CELY55F3AR/c
LOCUS       CELY55F3AR  14922 bp    DNA             INV       28-MAR-2000
DEFINITION  Caenorhabditis elegans cosmid Y55F3AR, complete sequence.
ACCESSION   AC024827
VERSION     AC024827.1  GI:7140389
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 14922)
  AUTHORS   The C. elegans Genome Sequencing Consortium, Washington University
            Genome Sequencing Center, St. Louis U.S.A. and the Sanger Centre,
            Hinxton, U.K.,C.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
  REMARK    Erratum:[[published errata appear in Science 1999 Jan
            1;283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep
            3;285(5433):1493]]
REFERENCE   2  (bases 1 to 14922)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 14922)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
                    Genome Sequencing Center
                    Department of Genetics, Washington University,
                    St. Louis, MO 63110, USA, and
                    Sanger Centre, Hinxton Hall
                    Cambridge CB10 IRQ, England
                    e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk


            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality
            >= 30); an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one m13 subclone.


            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
```

```
FEATURES             Location/Qualifiers
     source          1. .14922
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="IV"
                     /clone="Y55F3AR"
     gene            complement(1918. .8019)
                     /gene="Y55F3AR.3"
     CDS             complement(join(1918. .2108,3040. .3423,3987. .4221,
                     5321. .5427,5782. .6280,7546. .7620,7673. .7809,7872. .8019))
                     /gene="Y55F3AR.3"
                     /note="contains similarity to Pfam family PF00118
                     (TCP-1/cpn60 chaperonin family), score=441.7, E=6.5e-129,
                     N=2"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAF60806.1"
                     /db_xref="GI:7332119"
                     /translation="MAMKIPKSGYNRFMKECAQHFKGTDEAVQRNIEACTELASQIRS
                     AYGPNGMNKMVINHIEKLFVTNDAATILKELEIQHPAARIIIMATEMQEKQIGDNTNT
                     VVILAAALLEHAANLIHMGMTPQEVAAGYEQAAEKALEILPTLVVKEATDMKNIEEVR
                     QYIRSAITSKQYDNEDIIADLVAKACVTTCPANSFNFNVDNIRICKIIGSGVHTSTVM
                     NGMVFKRGAEGEIREARDARIAVYTCPFDLTQTETKGTVLIENADELRNFSKGEEAEV
                     EEQVKAIADNARSSILAKFWLMNFSKGEEAEVEEQQVKAIADNGVKVVVAAGKFGDMY
                     LHFLNKYKIMAVRLTSKFDLRRLCRTVGAQPQARICAPAVNLLGHCDSVAVQEIGDEN
                     VVVFDKKSASNLNFLKFSETGKVATIIIRGSSQSRIDDVERAVDDAVNTYKALTKDGK
                     LLAGAGAVEIELAKEIESFGAKAPGLEQYAIKKFAHALETLPKAIAENAGMPTTETLT
                     KLYAEHVAGKKNAGIDIWKRETMDAVVNNIFDLYAGKRLAIKLATDAASTILKVDQII
                     MSKQATGGPKPRGPKAQDEDDDGMA"
     gene            8656. .10604
                     /gene="Y55F3AR.2"
     CDS             join(8656. .8750,8813. .9068,9232. .9381,10191. .10604)
                     /gene="Y55F3AR.2"
                     /note="contains similarity to Pfam family PF00085
                     (Thioredoxin), score=120.2, E=6e-35, N=1; coded for by C.
                     elegans cDNA yk435d12.5; coded for by C. elegans cDNA
                     yk435d12.3"
                     /codon_start=1
                     /protein_id="AAF60805.1"
                     /db_xref="GI:7332118"
                     /translation="MPVIVVNGDSDFDRKFSAGNGKAVFVDFTASWCGPCQYIAPIFS
                     DLANQYKGSVFLKVDVDECRGTAATYGVNAMPTFIAFVNGQKKATIQGADESGLRSMV
                     AKYASTSAAWSGTGQPKSWQFSRFWFEKSQFSLKNYLFYVGLLEKTQNFSHFRQNQQK
                     KYFSSDFRLSGSTTGSSSSTSSSAPPRASGAAPPPAGPNPLEPVFRIIETINERLQGF
                     GISGVNFAGFQLQPFHLLLAIALLIIFGPFGLLIALAVCFFTQPRGGGPAPGAPRAGG
                     SGAGGAPRPPPPAGPRAFGGSGQRLGGN"
     gene            11568. .14837
                     /gene="Y55F3AR.1"
     CDS             join(11568. .11648,12053. .12310,14451. .14837)
                     /gene="Y55F3AR.1"
                     /note="contains similarity to human OXA1L (GB:AJ001981);
                     coded for by C. elegans cDNA yk647c9.5"
                     /codon_start=1
                     /protein_id="AAF60807.1"
                     /db_xref="GI:7332120"
                     /translation="MLSRICAPRHISIASTSFLNPQNRCETTISRRNMSALPPAIAPA
                     FQYASDCYIAQSIQFGMESLHGLGLSWPGAFISAAILLRIGTAPLHIYAEKLFANRLH
                     AQNFLTQGIMKKVSERYRVQLGPSTDGTKLEVKSSDPKIAKATQDALQEVPAMLAEHG
                     LQAARIQNLKMCTVPVWIFSSFALRNVINSDFHPSVAGHLWIPDMLAPDPYFILPVAV
                     GVFGFLNLYVRNLAFLDENLTKF"
BASE COUNT     4815 a    2553 c    2576 g    4978 t
ORIGIN


  Query Match             88.0%;  Score 17.6;  DB 6;  Length 14922;
  Best Local Similarity   85.0%;  Pred. No. 2.6e+02;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ttgstcaaaatttgascaaa 20
           |||:| |||||||||:||||
Db   11273 TTGGTGAAAATTTGAGCAAA 11254
```

```
RESULT   9
CELY19D10B/c
LOCUS       CELY19D10B  22964 bp    DNA             INV       24-MAR-2000
DEFINITION  Caenorhabditis elegans cosmid Y19D10B, complete sequence.
ACCESSION   AC006723
VERSION     AC006723.1  GI:4263215
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 22964)
  AUTHORS   The C. elegans Genome Sequencing Consortium, Washington University
            Genome Sequencing Center, St. Louis U.S.A. and the Sanger Centre,
            Hinxton, U.K.,C.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
  REMARK    Erratum:[[published errata appear in Science 1999 Jan
            1;283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep
            3;285(5433):1493]]
REFERENCE   2  (bases 1 to 22964)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 22964)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 22964)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
                    Genome Sequencing Center
                    Department of Genetics, Washington University,
                    St. Louis, MO 63110, USA, and
                    Sanger Centre, Hinxton Hall
                    Cambridge CB10 IRQ, England
                    e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk


            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.


            WARNING: These data have only had automated annotation
            and have not yet been subjected to manual review of that
            annotation.  We will be manually reviewing this information
            as quickly as possible and at that time this GenBank record
            will be updated and this warning removed.


            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES             Location/Qualifiers
     source          1. .22964
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="V"
                     /clone="CELY19D10B"
```

```
     gene            complement(221. .3399)
                     /gene="Y19D10B.f"
     CDS             complement(join(221. .619,668. .804,861. .1094,1143. .1288,
                     2248. .2386,2744. .3098,3145. .3281,3342. .3399))
                     /gene="Y19D10B.f"
                     /note="contains similarity to TR:O44597"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAF59429.1"
                     /db_xref="GI:7322372"
                     /translation="MSAAAKEPPMRQEASVRRPDDLDVMWALNNQMLELGLTNLTSIS
                     CAGFQVVSNKKLNKLNMPNLKNITLNEFPEPTDKAMKFSIDSGSPDFCITVEEMENFI
                     TADNATVDEIYGKYCPPTTTENLCTSPNICQKVFGNVEIGPNSNLDTMKSIEIIFGSL
                     IINETNLTDFNFLENLKYVAQLTKDKSAIIVQNNPFLLNISFPKLKVRFELPTVDQFY
                     YFSVCDPTISTPPFLTKTVLFSAVYFEFENLRMIGQCLVFLLLAESAFGENLTKFFQKK
                     NCDPKCTFETQNLTSHSLAFFPTDCSTVCTHLSINEHTDLTEAQLASTFLSMKRLIGS
                     LTVMDTKFKSISFLAGLEAVECGQYSIVLDFNNQMIEFGLTNLTSISCSNFEIIGYKK
                     LRNLGFPNLKSITLDEFSSPFDEFVRLYIDPDSIDFCITVEEMETLISAEDVNVDGIY
                     GKYCTPTTSENLCTFPEICPKIFGDVEVGSSSSIDSLKFMEIIFGSLIINGTNLTNFN
                     FLENLKYVAQLTSKFCGINVGTDQMG"
     gene            complement(3696. .4093)
                     /gene="Y19D10B.e"
     CDS             complement(join(3696. .3952,4006. .4093))
                     /gene="Y19D10B.e"
                     /note="contains similarity to TR:O44600"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAF59428.1"
                     /db_xref="GI:7322371"
                     /translation="MKEKCLVFLLCAVCAFCVNLTVFFSKKNCDPKCTFENQNATSKT
                     IGSFPTECSTVCSFLRITEDTDLTEAQLASTFLNMKRLLGGVIVISTKFKSGSFLAGL
                     EAIECGKQMRST"
     gene            complement(4582. .9384)
                     /gene="Y19D10B.d"
     CDS             complement(join(4582. .4650,4695. .4847,4904. .5013,
                     5057. .5131,5470. .5631,5678. .5743,5826. .5871,5921. .5983,
                     6028. .6150,6202. .6440,6487. .6634,6679. .6812,6935. .6976,
                     7020. .7179,7317. .7375,7424. .7697,7788. .8126,8174. .8393,
                     9162. .9267,9327. .9384))
                     /gene="Y19D10B.d"
                     /note="coded for by C. elegans cDNA yk19f1.5; coded for by
                     C. elegans cDNA yk28f3.5"
                     /codon_start=1
                     /product="Hypothetical protein Y19D10B.d"
                     /protein_id="AAF59427.1"
                     /db_xref="GI:7322370"
                     /translation="MKTRHYSLLIFLAGSAFFPEISAQLANCQNLGTPINSTDCLCPG
                     FVDGRLCENIKCQRWAVPDKNRCACAPGWYDRYCGLRGCRPPNEEQMDLEKRSFIIVF
                     NTKTTMKMQLDSLKNNFKEMISKVSKNSKGVVNNWIENYIVYGFVESQSHTKIQKLFT
                     ENPDDVISFFNNMTLFDGDDTQPVLTAIQNAQQTYPKMQAHALVLVFTDSPASDATAW
                     SHRFTDKNAEQSVLQISLIWRSKYSFFLSLPARTNFLSNGVDVYRRLSLTNHGDTFFI
                     KDSNDLSNALLNVIGVQYFPENVQIGYGKTTDEQLQTYVDNDGDLVYILLTIRSSLQT
                     TVPSISEATIVAEGDSYRLYSPASNIGDTVTIHCQNGTTYNYRMFIQSKNTLLFNYND
                     DMHIDVGNGMAVIGAKSIFLGTKNSANRDHNARDLSGNANELTTNVSKNNTDTCSKKN
                     LDSINDPRLQEPRQFIFILEQHSGNEQVYKTLAREINDILELVNSSTTSEYKKEFTLI
                     AHDSAESHVLFSSYNPKIFAEKFSKLVNSLELTDNLDNTMGLLSIIHAQKMNILPTAL
                     VYYFTNQAVKNAQNLTRNWDLVKRDIEINFLTIADGVTTEIFALPKQLELVQKMTNGR
                     LIPLGKTENSIFPIFKDMMTVTTLTTDNEQYNCHDTALEIDGYFEDGASVSISNYVTY
                     KNPNFMSIRIDSKLFASGKWKLSALSTTGGCQITVRQKTSVGLVLGFTAATNDDNAST
                     QIISQRSSSDSQPIFTTLKTTNHIVPTDLEIQIVNRQRYDQPVSYVKLGNAQRDTNSC
                     SYDFISDSVVVPKSTLTMWIVSATDNAGKLILHRIFYYYLPRILSEFAVVLQGFMEIF
                     VKLCSCSPPTLLQCLQYRPLQSQHK"
     gene            12528. .15727
                     /gene="Y19D10B.g"
     CDS             join(12528. .12585,12627. .12715,12795. .13068,13117. .13175,
                     13362. .13521,13564. .13756,13799. .13866,13910. .14013,
                     15623. .15727)
                     /gene="Y19D10B.g"
                     /note="coded for by C. elegans cDNA yk19f1.5; coded for by
                     C. elegans cDNA yk28f3.5"
                     /codon_start=1
                     /product="Hypothetical protein Y19D10B.g"
                     /protein_id="AAF59430.1"
                     /db_xref="GI:7322373"
                     /translation="MTHNPYLLPSRMHNRRIRKSSDATAWSHRFTDKNAEQSVLQISL
                     IWRSKYSIFLSLPARTNFSSNGVDVYRRLSLTNHGDTFFIKDSNDLSNALLIVIGVQY
                     FPENVQIGYGKTTDEQLQTYVDNDGDLVYILLTIRSSLQATVPSISEATIVAEGDSYR
                     LYSPASNIGDTVTILCQNGTRYNYRMFIQSKNTLLFNYNDDMHIDVGNGMAVIGIEMF
                     ATMQTYGFPQWQNSSYDVRSSDGKLLREKFYASTRPQQDCTFLYGFPAWKTLSCPPGP
                     IMQLHTFYYNGYNQQRITPGYCIESDHNARDLSGNAETLNELTINVSKINMDTLSKKN
                     AIGAYCTTVCGEPSDGLAVDVILVENALVFCKTYM"
     gene            complement(12647. .18655)
                     /gene="Y19D10B.c"
     CDS             complement(join(12647. .12741,14320. .14464,14511. .14948,
                     15001. .15120,15428. .15708,16816. .16856,17136. .17255,
                     17308. .17363,17815. .18036,18082. .18238,18600. .18655))
                     /gene="Y19D10B.c"
                     /note="contains similarity to TR:O44603"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAF59426.1"
                     /db_xref="GI:7322369"
                     /translation="MAPPHSTMVSPRASKINGWKTLLNCLQTIKCIQNVSLDPCNDFY
                     DYVCTNDTRIITKLNFTSLFEDREILVNPNISFPNNGTYITLDEILDPDSLMIWYLKI
                     ELLNTTTADFDYFFKMHREYVKKENKSEKFSFVVETLENVNVTIIITKSNMISNDLAD
                     SVGRPQYFPSKKICLFWVMLSWFLRRLQYVESDESTLLLFANTMSYYQASKYLSTSIK
                     HQCIFYEYYINSETVTWFPTNCSTVCADLKIDRICDLTEDQLKSTFRNMKHLIGSLTV
                     VNTNHKSGEFLDGLESLECGGEFRKRRTATDAPISQFVWNANSNMTEIGLINFTSIAC
                     LSMEIFNVKMLNMPKLKNFTNPFGNDMKIGITIMVPDRYFCISPQEMLNFLSSDNVLT
                     PEINGKYCDLPSSVSNGTVCNSTSLTDGCTQIFGSLVIGPENEQFVSQLKYVEMIFGR
                     LVVNNTNLTNIDFLQSLKYIYSLDGSTPAISVENNPFLSNFSFPNLKLAKTKGRSTIW
                     FNKNSEALSSDSAYCYKLQKSLNLNENQLIYFDEKTCAFMKISKTYFRSPYKRDLEYT
                     LFGIFICESM"
     gene            complement(18918. .21277)
                     /gene="Y19D10B.b"
     CDS             complement(join(18918. .19055,19106. .19282,19454. .19567,
                     19627. .19773,19898. .20108,20152. .20453,20540. .20761,
                     20810. .20920,21221. .21277))
                     /gene="Y19D10B.b"
                     /note="contains similarity to TR:O44603"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAF59425.1"
                     /db_xref="GI:7322368"
                     /translation="MNFLLILTLISIIVRSRQENASMDPCNDFYDYVCTNDTRSITKL
                     NFASLFDDREILKPNISFPNNGTYIALDQILEPDRLMKWYMTVKSLDTATVDLEYFFE
                     THYEYISKKNDSEKFSFAVETLENVNLTLTNIFYNSLYANIKVLNQLNLPAAEENKFL
                     HHISNLLKNNTIEEIWSSSLSNKSKKVLDEELNNSKVFFGLLDYNNVTMLEETKLVYE
                     TEYYRLKSLLSSDDLDTEIAEKILRLGAIDTATKYLAKRIPGYEPEFLFSAFASNMEN
                     EAFKHNSAIYAGIVTRSDLQQPMLAIADTVSVLAHELMHFVYPSGTELLTPNVTKAAK
                     KCTQDEVKRMGDSDVVKPIDGWFSKKVTHEDLANIMGLRMVMKMVARKSASEEQLKNA
                     LETLLSGLCIQGKPKNSILPHHHPFEISINTAVRQYPLFSSLYGCRLGDRMFAEAEQF
                     CKPLGGEVNLEDYKIKNDTEDIGGFFTYIMDTANAFNFSFRN"
     gene            complement(22263. .22764)
                     /gene="Y19D10B.a"
     CDS             complement(join(22263. .22462,22509. .22764))
                     /gene="Y19D10B.a"
                     /note="contains similarity to TR:O44604"
                     /codon_start=1
                     /evidence=not_experimental

  Query Match             88.0%;  Score 17.6;  DB 6;  Length 22964;
  Best Local Similarity   85.0%;  Pred. No. 2.6e+02;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ttgstcaaaatttgascaaa 20
           |||:| |||||||||:||||
Db    1184 TTGGTGAAAATTTGACCAAA 1165


RESULT 10
CEC54C6
LOCUS       CEC54C6     35500 bp    DNA             INV       09-NOV-2000
DEFINITION  Caenorhabditis elegans cosmid C54C6, complete sequence.
ACCESSION   Z77131
VERSION     Z77131.1  GI:1438693
```

KEYWORDS HTG; 60S ribosomal protein L37; Tubulin.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 35500)
AUTHORS none.
TITLE Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
REMARK The C.elegans Sequencing Consortium.
Erratum:[[published errata appear in Science 1999 Jan 1;283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep 3;285(5433):1493]]
REFERENCE 2 (bases 1 to 35500)
AUTHORS Lightning,J.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-1996) Nematode Sequencing Project, Sanger Centre, Hinxton, Cambridge CB10 1RQ, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.
Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is not the entire insert of clone C54C6. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true left end of clone C54C6 is at 1 in this sequence. The true right end of clone C54C6 is at 6823 in sequence Z36237.
The true left end of clone C48D5 is at 35399 in this sequence. The start of this sequence (1. .104) overlaps with the end of sequence Z46381.
The end of this sequence (35399. .35500) overlaps with the start of sequence Z36237.
For a graphical representation of this sequence and its analysis see:- http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=C54C6.

FEATURES Location/Qualifiers

source 1. .35500
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="III"
/clone="C54C6"

gene 5549. .5824
/gene="rpl-37"
/note="C54C6.1"

CDS 5549. .5824
/gene="rpl-37"
/note="predicted using Genefinder
Similarity to putative C.elegans 60S ribosomal protein L37 (SW:RL37_CAEEL), contains similarity to Pfam domain: PF01907 (Ribosomal protein L37e), Score=127.1, E-value=1.1e-34, N=1
cDNA EST CEMSB96R comes from this gene
cDNA EST CEMSD31F comes from this gene
cDNA EST CEMSE32F comes from this gene
cDNA EST yk494g7.5 comes from this gene
cDNA EST yk495b10.5 comes from this gene
cDNA EST yk495h10.5 comes from this gene
cDNA EST yk535a9.5 comes from this gene
cDNA EST yk535c2.5 comes from this gene
cDNA EST yk538a4.5 comes from this gene
cDNA EST yk540d1.5 comes from this gene
cDNA EST yk551h3.5 comes from this gene
cDNA EST yk558e8.5 comes from this gene
cDNA EST yk564c3.5 comes from this gene
cDNA EST yk578b8.5 comes from this gene
cDNA EST yk651b9.5 comes from this gene
cDNA EST yk668h9.5 comes from this gene"
/codon_start=1
/protein_id="CAB00854.1"
/db_xref="GI:3875183"
/db_xref="SWISS-PROT:P49622"
/translation="MTKGTQAFGKKHVKSHTLCKRCGKSSFHIQKKRCASCGYQDAKK
RTYNWGAKSIRRRTTGTGRTRHLRDVNARFRNGFRGTTPKPRAQPTN"

gene complement(join(15233. .15408,16194. .16691,16749. .16911, 16964. .17319,18335. .18500))
/gene="C54C6.2"
/note="ben-1"

CDS complement(join(15233. .15408,16194. .16691,16749. .16911, 16964. .17319,18335. .18500))
/gene="C54C6.2"
/note="predicted using Genefinder
contains similarity to Pfam domain: PF00091 (Tubulin/FtsZ family), Score=959.4, E-value=3.1e-285, N=1
cDNA EST yk119d7.5 comes from this gene
cDNA EST yk119d7.3 comes from this gene"
/codon_start=1
/protein_id="CAB00853.1"
/db_xref="GI:3875182"
/db_xref="SPTREMBL:Q18817"
/translation="MREIVHVQAGQCGNQIGAKFWEVISDEHGIQPDGTYKGESDLQL
ERINVYYNEANGGKYVPRAVLVDLEPGTMDSVRSGPFGQLFRPDNFVFGQSGAGNNWA
KGHYTEGAELVDNVLDVVRKEAEGCDCLQGFQLTHSLGGGTGSGMGTLLISKIREEYP
DRIMSSFSVVPSPKVSDTVVEPYNATLSVHQLVENTDETFCIDNEALYDICFRTLKLS
NPTYGDLNHLVSVTMSGVTTCLRFPGQLNADLRKLAVNMVPFPRLHFFMPGFAPLSAK
GAQAYRALTVAELTQQMFDAKNMMAACDPRHGRYLTVAAMFRGRMSMREVDDQMMNVQ
NKNSSYFVEWIPNNVKTAVCDIPPRGLKMSATFIGNSTAIQELFKRISEQFTGTIMET
RKTMFRRKAFLHWYTGEGMDEMEFTEAESNMNDLVSEYQQYQEATAEEDGELDGTDGD
AE"

gene complement(join(26246. .26385,26469. .26745))
/gene="C54C6.5"

CDS complement(join(26246. .26385,26469. .26745))
/gene="C54C6.5"
/codon_start=1
/protein_id="CAB00856.1"
/db_xref="GI:3875185"
/db_xref="SPTREMBL:Q18819"
/translation="MKCSIFLALVLFSAVLLFQSEASSVPSSEVQNTLFKPSSVVFRA
KCKMMCSQMCSPVLVSKKMMTCIHISMNIFREYVIGTTNAFVSIIIAFFKIMTEHIFL
SKMTENKKTKKKKGFFFQNTTTIKKFGNYFWLVIFK"

gene join(31992. .32092,33022. .33391)
/gene="C54C6.4"

CDS join(31992. .32092,33022. .33391)
/gene="C54C6.4"
/codon_start=1
/protein_id="CAB00855.1"
/db_xref="GI:3875184"
/db_xref="SPTREMBL:Q18818"
/translation="MLFFVNRSLVKKSSASKCNLSERYQLTENIKALRLLIAKDHFGN
ISKNIYIKFVTFQTIPAVFYGNIRKLYGFVYEHVIIPYRYRIQLSNMPKSPQLCSNLS
LHYFCMSNIRMTFSFVYFCFFCFRLLQLSTWLLLSTYFIIIVKRGCSKKLELRR"

CDS join(complement(6557. .6634),complement(4620. .4752), complement(3026. .3516),complement(2039. .2460), complement(1600. .1993),complement(503. .599), complement(332. .455),complement(105. .281), complement(Z46381.1:36211. .36355), complement(Z46381.1:35169. .35886), complement(Z46381.1:35020. .35113), complement(Z46381.1:34659. .34728))
/gene="M01F1.7"
/note="contains similarity to Pfam domain: PF02121 (Phosphatidylinositol transfer protein), Score=666.4, E-value=4.8e-197, N=1
cDNA EST yk7g1.5 comes from this gene
cDNA EST yk278g1.3 comes from this gene
cDNA EST yk278g1.5 comes from this gene"
/codon_start=1

```
                     /protein_id="CAB00857.1"
                     /db_xref="GI:3875186"
                     /db_xref="SPTREMBL:O62119"
                     /translation="MLIKEYRILLPMTVQEYRIAQLYMIQKKSRLDSHGQDSGVEIIS
                     NKPYTDGPGGSGQYTFKIYHIGSRIPAWIRTVLPTNALEAHEESWNAYPVTKTRYSTP
                     MMDRFSLEVETLYFDDHGQQENVFNLNEKDKSTRIIDYMDFVKDPISSHDYCAEEDPK
                     LYRSETTNRGPLNDDWVAEHLKKGLPIMCAYKLCKVEFRYWGMQTRAERWIHDLALRN
                     TMMRAHRQAWAWQDEWTGLTMNDIRKLEAEAALHLSKVMSVKENEDGHQDENDTDDDM
                     DAGDAVSDDLYFDCTDTSPIPTQKPSIIRWSSELELEIQDDSEYQNSEKFVKTDKILD
                     SPPLTPHNGSTEVALLIMVFHGDFSPDNPADSKTTDTNTFSSTIETCVQRHYPQLRNR
                     LHIVNVSCGHEMTQVVSKLSNISPSFGLLHPSLSLMLPSASHLYNEAVEGTIRRANET
                     YNEFIASQPLFNGEVFVVGDCVGGIFLYEAMTRKCDSMTLLKRLSSNLSSRIIKEDFG
                     ASRLSFQPSTVFLLGCPLGLTLMQRKLDGSEIDGLDSCQLFNLYYPLDPCGARIEPVL
                     DGQLSCVPPYNVPKYQRYPLGDGKSQKFESTIDATQMWGSKRIDNLLYCPNSMVALPS
                     SALPNILHASYWESCDVASFLLRQFVRGEENSVLTTLSSSMNNIPLNIDLPTMHWKRK
                     RTRFKIANLSANHRANDILVTAGMDLTVIAKFCYGPMDLVALSREPVSVFVYPQRGDW
                     YLHGVFDTDSHGRLTLQLAKTLPCGIHSVKIVVHGDRSYLDAFVAIVPHGTKCAVFSV
                     DGSLTASVSVTGKDPRVRPGAVDVVRYWQEQGYLIIYLTARPDMQQRVVSAWLAQHNF
                     PHALLFFNNSFSTEPLKQKSLHLRHIVDMGVHIHVAYGSGKDVNVYTSAGVDPEHVIS
                     VAGSRRRNCVQIESYSSHLAALNSGQCTLGKRIEDDGLTLQLHRNVQRTPSFTPRGGK
                     FENEKDRR"
     gene            join(complement(6557. .6634),complement(4620. .4752),
                     complement(3026. .3516),complement(2039. .2460),
                     complement(1600. .1993),complement(503. .599),
                     complement(332. .455),complement(105. .281),
                     complement(Z46381.1:36211. .36355),
                     complement(Z46381.1:35169. .35886),
                     complement(Z46381.1:35020. .35113),
                     complement(Z46381.1:34659. .34728))
                     /gene="M01F1.7"
BASE COUNT    11940 a   5924 c   6270 g  11366 t
ORIGIN


  Query Match             88.0%;  Score 17.6;  DB 5;  Length 35500;
  Best Local Similarity   85.0%;  Pred. No. 2.6e+02;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||:|||||||||||: |||
Db   5438 TTGGTCAAAATTTGACAAAA 5457


RESULT  11
AC012726
LOCUS       AC012726    42728 bp    DNA             HTG       03-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC012726
VERSION     AC012726.1  GI:6223581
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 42728)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10209813 by the submitter.
            For further information on this sequence you may e-mail to
            fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .42728
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    13034 a   8825 c   8764 g  12105 t
ORIGIN


  Query Match             88.0%;  Score 17.6;  DB 63;  Length 42728;
  Best Local Similarity   85.0%;  Pred. No. 2.6e+02;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||:|||| ||||||:||||
Db   3676 TTGGTCAATATTTGACCAAA 3695


RESULT  12
AC009745
LOCUS       AC009745    71585 bp    DNA             HTG       25-NOV-1999
DEFINITION  Drosophila melanogaster chromosome 2 clone BACR01H19 (D1039)
            RPCI-98 01.H.19 map 39D-39D strain y; cn bw sp, *** SEQUENCING IN
            PROGRESS ***, 77 unordered pieces.
ACCESSION   AC009745
VERSION     AC009745.7  GI:6468061
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 71585)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71585)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Nov 25, 1999 this sequence version replaced gi:6136347.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 77 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1      390: contig of 390 bp in length
            *     391      470: gap of unknown length
            *     471      888: contig of 418 bp in length
            *     889      968: gap of unknown length
            *     969     1506: contig of 538 bp in length
            *    1507     1586: gap of unknown length
            *    1587     2604: contig of 1018 bp in length
            *    2605     2684: gap of unknown length
            *    2685     3574: contig of 890 bp in length
```

```
*     3575     3654: gap of unknown length
*     3655     4159: contig of 505 bp in length
*     4160     4239: gap of unknown length
*     4240     9257: contig of 5018 bp in length
*     9258     9337: gap of unknown length
*     9338    17468: contig of 8131 bp in length
*    17469    17548: gap of unknown length
*    17549    24956: contig of 7408 bp in length
*    24957    25036: gap of unknown length
*    25037    25571: contig of 535 bp in length
*    25572    25651: gap of unknown length
*    25652    26012: contig of 361 bp in length
*    26013    26092: gap of unknown length
*    26093    26726: contig of 634 bp in length
*    26727    26806: gap of unknown length
*    26807    27446: contig of 640 bp in length
*    27447    27526: gap of unknown length
*    27527    28267: contig of 741 bp in length
*    28268    28347: gap of unknown length
*    28348    28931: contig of 584 bp in length
*    28932    29011: gap of unknown length
*    29012    29803: contig of 792 bp in length
*    29804    29883: gap of unknown length
*    29884    30825: contig of 942 bp in length
*    30826    30905: gap of unknown length
*    30906    31261: contig of 356 bp in length
*    31262    31341: gap of unknown length
*    31342    31640: contig of 299 bp in length
*    31641    31720: gap of unknown length
*    31721    32811: contig of 1091 bp in length
*    32812    32891: gap of unknown length
*    32892    33709: contig of 818 bp in length
*    33710    33789: gap of unknown length
*    33790    34402: contig of 613 bp in length
*    34403    34482: gap of unknown length
*    34483    35218: contig of 736 bp in length
*    35219    35298: gap of unknown length
*    35299    36114: contig of 816 bp in length
*    36115    36194: gap of unknown length
*    36195    36696: contig of 502 bp in length
*    36697    36776: gap of unknown length
*    36777    38114: contig of 1338 bp in length
*    38115    38194: gap of unknown length
*    38195    38656: contig of 462 bp in length
*    38657    38736: gap of unknown length
*    38737    39478: contig of 742 bp in length
*    39479    39558: gap of unknown length
*    39559    39859: contig of 301 bp in length
*    39860    39939: gap of unknown length
*    39940    40418: contig of 479 bp in length
*    40419    40498: gap of unknown length
*    40499    40882: contig of 384 bp in length
*    40883    40962: gap of unknown length
*    40963    41585: contig of 623 bp in length
*    41586    41665: gap of unknown length
*    41666    42346: contig of 681 bp in length
*    42347    42426: gap of unknown length
*    42427    42911: contig of 485 bp in length
*    42912    42991: gap of unknown length
*    42992    43200: contig of 209 bp in length
*    43201    43280: gap of unknown length
*    43281    43737: contig of 457 bp in length
*    43738    43817: gap of unknown length
*    43818    44347: contig of 530 bp in length
*    44348    44427: gap of unknown length
*    44428    45684: contig of 1257 bp in length
*    45685    45764: gap of unknown length
*    45765    46377: contig of 613 bp in length
*    46378    46457: gap of unknown length
*    46458    47060: contig of 603 bp in length
*    47061    47140: gap of unknown length
*    47141    47737: contig of 597 bp in length
*    47738    47817: gap of unknown length
*    47818    48438: contig of 621 bp in length
*    48439    48518: gap of unknown length
*    48519    48757: contig of 239 bp in length
*    48758    48837: gap of unknown length
*    48838    49385: contig of 548 bp in length
*    49386    49465: gap of unknown length
*    49466    50193: contig of 728 bp in length
*    50194    50273: gap of unknown length
*    50274    50839: contig of 566 bp in length
*    50840    50919: gap of unknown length
*    50920    51360: contig of 441 bp in length
*    51361    51440: gap of unknown length
*    51441    52204: contig of 764 bp in length
*    52205    52284: gap of unknown length
*    52285    52817: contig of 533 bp in length
*    52818    52897: gap of unknown length
*    52898    53632: contig of 735 bp in length
*    53633    53712: gap of unknown length
*    53713    54299: contig of 587 bp in length
*    54300    54379: gap of unknown length
*    54380    54999: contig of 620 bp in length
*    55000    55079: gap of unknown length
*    55080    55855: contig of 776 bp in length
*    55856    55935: gap of unknown length
*    55936    56610: contig of 675 bp in length
*    56611    56690: gap of unknown length
*    56691    57519: contig of 829 bp in length
*    57520    57599: gap of unknown length
*    57600    58236: contig of 637 bp in length
*    58237    58316: gap of unknown length
*    58317    59016: contig of 700 bp in length
*    59017    59096: gap of unknown length
*    59097    59662: contig of 566 bp in length
*    59663    59742: gap of unknown length
*    59743    60327: contig of 585 bp in length
*    60328    60407: gap of unknown length
*    60408    61195: contig of 788 bp in length
*    61196    61275: gap of unknown length
*    61276    61948: contig of 673 bp in length
*    61949    62028: gap of unknown length
*    62029    62609: contig of 581 bp in length
*    62610    62689: gap of unknown length
*    62690    63036: contig of 347 bp in length
*    63037    63116: gap of unknown length
*    63117    63606: contig of 490 bp in length
*    63607    63686: gap of unknown length
*    63687    64621: contig of 935 bp in length
*    64622    64701: gap of unknown length
*    64702    65145: contig of 444 bp in length
*    65146    65225: gap of unknown length
*    65226    65824: contig of 599 bp in length
*    65825    65904: gap of unknown length
*    65905    66506: contig of 602 bp in length
*    66507    66586: gap of unknown length
*    66587    67066: contig of 480 bp in length
*    67067    67146: gap of unknown length
*    67147    67564: contig of 418 bp in length
*    67565    67644: gap of unknown length
*    67645    68050: contig of 406 bp in length
*    68051    68130: gap of unknown length
*    68131    68846: contig of 716 bp in length
*    68847    68926: gap of unknown length
*    68927    69136: contig of 210 bp in length
*    69137    69216: gap of unknown length
*    69217    69732: contig of 516 bp in length
*    69733    69812: gap of unknown length
*    69813    70251: contig of 439 bp in length
*    70252    70331: gap of unknown length

Query Match             88.0%;  Score 17.6;  DB 61;  Length 71585;
Best Local Similarity   85.0%;  Pred. No. 2.5e+02;
Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;
```

```
Qy       1 ttgstcaaaatttgascaaa 20
           |||:||||||||||: |||
Db    5714 TTGCTCAAAATTTGAGAAAA 5733


RESULT 13
AC008307/c
LOCUS       AC008307    92237 bp    DNA             HTG       25-FEB-2000
DEFINITION  Drosophila melanogaster chromosome 3 clone BACR03D22 (D709) RPCI-98
            03.D.22 map 86F-87A strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 92 unordered pieces.
ACCESSION   AC008307
VERSION     AC008307.6  GI:7105476
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 92237)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 92237)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Feb 25, 2000 this sequence version replaced gi:7021558.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 92 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      627: contig of 627 bp in length
            *      628      707: gap of unknown length
            *      708     1382: contig of 675 bp in length
            *     1383     1462: gap of unknown length
            *     1463     2164: contig of 702 bp in length
            *     2165     2244: gap of unknown length
            *     2245     2829: contig of 585 bp in length
            *     2830     2909: gap of unknown length
            *     2910     3640: contig of 731 bp in length
            *     3641     3720: gap of unknown length
            *     3721     4562: contig of 842 bp in length
            *     4563     4642: gap of unknown length
            *     4643     5305: contig of 663 bp in length
            *     5306     5385: gap of unknown length
            *     5386     6178: contig of 793 bp in length
            *     6179     6258: gap of unknown length
            *     6259     7195: contig of 937 bp in length
            *     7196     7275: gap of unknown length
            *     7276     7900: contig of 625 bp in length
            *     7901     7980: gap of unknown length
            *     7981     8893: contig of 913 bp in length
            *     8894     8973: gap of unknown length
            *     8974     9811: contig of 838 bp in length
            *     9812     9891: gap of unknown length
            *     9892    10540: contig of 649 bp in length
            *    10541    10620: gap of unknown length
            *    10621    11367: contig of 747 bp in length
            *    11368    11447: gap of unknown length
            *    11448    11976: contig of 529 bp in length
            *    11977    12056: gap of unknown length
            *    12057    12824: contig of 768 bp in length
            *    12825    12904: gap of unknown length
            *    12905    13991: contig of 1087 bp in length
            *    13992    14071: gap of unknown length
            *    14072    14955: contig of 884 bp in length
            *    14956    15035: gap of unknown length
            *    15036    16195: contig of 1160 bp in length
            *    16196    16275: gap of unknown length
            *    16276    17235: contig of 960 bp in length
            *    17236    17315: gap of unknown length
            *    17316    18235: contig of 920 bp in length
            *    18236    18315: gap of unknown length
            *    18316    19340: contig of 1025 bp in length
            *    19341    19420: gap of unknown length
            *    19421    20723: contig of 1303 bp in length
            *    20724    20803: gap of unknown length
            *    20804    21941: contig of 1138 bp in length
            *    21942    22021: gap of unknown length
            *    22022    22599: contig of 578 bp in length
            *    22600    22679: gap of unknown length
            *    22680    23187: contig of 508 bp in length
            *    23188    23267: gap of unknown length
            *    23268    24639: contig of 1372 bp in length
            *    24640    24719: gap of unknown length
            *    24720    25644: contig of 925 bp in length
            *    25645    25724: gap of unknown length
            *    25725    26747: contig of 1023 bp in length
            *    26748    26827: gap of unknown length
            *    26828    27449: contig of 622 bp in length
            *    27450    27529: gap of unknown length
            *    27530    28590: contig of 1061 bp in length
            *    28591    28670: gap of unknown length
            *    28671    29341: contig of 671 bp in length
            *    29342    29421: gap of unknown length
            *    29422    30147: contig of 726 bp in length
            *    30148    30227: gap of unknown length
            *    30228    31012: contig of 785 bp in length
            *    31013    31092: gap of unknown length
            *    31093    32304: contig of 1212 bp in length
            *    32305    32384: gap of unknown length
            *    32385    34041: contig of 1657 bp in length
            *    34042    34121: gap of unknown length
            *    34122    34937: contig of 816 bp in length
            *    34938    35017: gap of unknown length
            *    35018    35633: contig of 616 bp in length
            *    35634    35713: gap of unknown length
            *    35714    36852: contig of 1139 bp in length
            *    36853    36932: gap of unknown length
            *    36933    38147: contig of 1215 bp in length
            *    38148    38227: gap of unknown length
            *    38228    39797: contig of 1570 bp in length
            *    39798    39877: gap of unknown length
            *    39878    40925: contig of 1048 bp in length
            *    40926    41005: gap of unknown length
            *    41006    42786: contig of 1781 bp in length
            *    42787    42866: gap of unknown length
            *    42867    44129: contig of 1263 bp in length
            *    44130    44209: gap of unknown length
            *    44210    44946: contig of 737 bp in length
```

```
          *    44947    45026: gap of unknown length
          *    45027    46653: contig of 1627 bp in length
          *    46654    46733: gap of unknown length
          *    46734    48522: contig of 1789 bp in length
          *    48523    48602: gap of unknown length
          *    48603    50683: contig of 2081 bp in length
          *    50684    50763: gap of unknown length
          *    50764    53999: contig of 3236 bp in length
          *    54000    54079: gap of unknown length
          *    54080    56251: contig of 2172 bp in length
          *    56252    56331: gap of unknown length
          *    56332    59477: contig of 3146 bp in length
          *    59478    59557: gap of unknown length
          *    59558    64282: contig of 4725 bp in length
          *    64283    64362: gap of unknown length
          *    64363    64956: contig of 594 bp in length
          *    64957    65036: gap of unknown length
          *    65037    65812: contig of 776 bp in length
          *    65813    65892: gap of unknown length
          *    65893    66211: contig of 319 bp in length
          *    66212    66291: gap of unknown length
          *    66292    66632: contig of 341 bp in length
          *    66633    66712: gap of unknown length
          *    66713    67455: contig of 743 bp in length
          *    67456    67535: gap of unknown length
          *    67536    68305: contig of 770 bp in length
          *    68306    68385: gap of unknown length
          *    68386    69096: contig of 711 bp in length
          *    69097    69176: gap of unknown length
          *    69177    69775: contig of 599 bp in length
          *    69776    69855: gap of unknown length
          *    69856    70617: contig of 762 bp in length
          *    70618    70697: gap of unknown length
          *    70698    71259: contig of 562 bp in length
          *    71260    71339: gap of unknown length
          *    71340    71946: contig of 607 bp in length
          *    71947    72026: gap of unknown length
          *    72027    72791: contig of 765 bp in length
          *    72792    72871: gap of unknown length
          *    72872    73199: contig of 328 bp in length
          *    73200    73279: gap of unknown length
          *    73280    73960: contig of 681 bp in length
          *    73961    74040: gap of unknown length
          *    74041    74779: contig of 739 bp in length
          *    74780    74859: gap of unknown length
          *    74860    75203: contig of 344 bp in length
          *    75204    75283: gap of unknown length
          *    75284    76068: contig of 785 bp in length
          *    76069    76148: gap of unknown length
          *    76149    76573: contig of 425 bp in length
          *    76574    76653: gap of unknown length
          *    76654    76998: contig of 345 bp in length
          *    76999    77078: gap of unknown length
          *    77079    77664: contig of 586 bp in length
          *    77665    77744: gap of unknown length
          *    77745    78075: contig of 331 bp in length
          *    78076    78155: gap of unknown length
          *    78156    78872: contig of 717 bp in length
          *    78873    78952: gap of unknown length
          *    78953    79694: contig of 742 bp in length
          *    79695    79774: gap of unknown length

  Query Match             88.0%;  Score 17.6;  DB 60;  Length 92237;
  Best Local Similarity   85.0%;  Pred. No. 2.5e+02;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ttgstcaaaatttgascaaa 20
           |||:|||||||| |:||||
Db   83875 TTGGTCAAAATTTAACCAAA 83856


RESULT  14
AC009254
```

```
LOCUS       AC009254    95668 bp    DNA             HTG       20-MAR-2000
DEFINITION  Drosophila melanogaster chromosome 2 clone BACR42E05 (D1040)
            RPCI-98 42.E.5 map 40A-40C strain y; cn bw sp, *** SEQUENCING IN
            PROGRESS ***, 71 unordered pieces.
ACCESSION   AC009254
VERSION     AC009254.7  GI:7264756
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 95668)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 95668)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Mar 20, 2000 this sequence version replaced gi:6980147.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 71 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      901: contig of 901 bp in length
            *      902      981: gap of unknown length
            *      982     2063: contig of 1082 bp in length
            *     2064     2143: gap of unknown length
            *     2144     2761: contig of 618 bp in length
            *     2762     2841: gap of unknown length
            *     2842     3926: contig of 1085 bp in length
            *     3927     4006: gap of unknown length
            *     4007     4361: contig of 355 bp in length
            *     4362     4441: gap of unknown length
            *     4442     5512: contig of 1071 bp in length
            *     5513     5592: gap of unknown length
            *     5593     6722: contig of 1130 bp in length
            *     6723     6802: gap of unknown length
            *     6803     7344: contig of 542 bp in length
            *     7345     7424: gap of unknown length
            *     7425     8672: contig of 1248 bp in length
            *     8673     8752: gap of unknown length
            *     8753     9699: contig of 947 bp in length
            *     9700     9779: gap of unknown length
            *     9780    10657: contig of 878 bp in length
            *    10658    10737: gap of unknown length
            *    10738    11423: contig of 686 bp in length
```

```
*    11424    11503: gap of unknown length
*    11504    12759: contig of 1256 bp in length
*    12760    12839: gap of unknown length
*    12840    14234: contig of 1395 bp in length
*    14235    14314: gap of unknown length
*    14315    15453: contig of 1139 bp in length
*    15454    15533: gap of unknown length
*    15534    16474: contig of 941 bp in length
*    16475    16554: gap of unknown length
*    16555    17564: contig of 1010 bp in length
*    17565    17644: gap of unknown length
*    17645    19120: contig of 1476 bp in length
*    19121    19200: gap of unknown length
*    19201    20747: contig of 1547 bp in length
*    20748    20827: gap of unknown length
*    20828    21771: contig of 944 bp in length
*    21772    21851: gap of unknown length
*    21852    23126: contig of 1275 bp in length
*    23127    23206: gap of unknown length
*    23207    24497: contig of 1291 bp in length
*    24498    24577: gap of unknown length
*    24578    25753: contig of 1176 bp in length
*    25754    25833: gap of unknown length
*    25834    28163: contig of 2330 bp in length
*    28164    28243: gap of unknown length
*    28244    30342: contig of 2099 bp in length
*    30343    30422: gap of unknown length
*    30423    31251: contig of 829 bp in length
*    31252    31331: gap of unknown length
*    31332    32749: contig of 1418 bp in length
*    32750    32829: gap of unknown length
*    32830    34444: contig of 1615 bp in length
*    34445    34524: gap of unknown length
*    34525    36523: contig of 1999 bp in length
*    36524    36603: gap of unknown length
*    36604    38039: contig of 1436 bp in length
*    38040    38119: gap of unknown length
*    38120    39981: contig of 1862 bp in length
*    39982    40061: gap of unknown length
*    40062    42624: contig of 2563 bp in length
*    42625    42704: gap of unknown length
*    42705    44841: contig of 2137 bp in length
*    44842    44921: gap of unknown length
*    44922    47186: contig of 2265 bp in length
*    47187    47266: gap of unknown length
*    47267    51551: contig of 4285 bp in length
*    51552    51631: gap of unknown length
*    51632    55048: contig of 3417 bp in length
*    55049    55128: gap of unknown length
*    55129    60695: contig of 5567 bp in length
*    60696    60775: gap of unknown length
*    60776    65860: contig of 5085 bp in length
*    65861    65940: gap of unknown length
*    65941    71090: contig of 5150 bp in length
*    71091    71170: gap of unknown length
*    71171    76900: contig of 5730 bp in length
*    76901    76980: gap of unknown length
*    76981    77362: contig of 382 bp in length
*    77363    77442: gap of unknown length
*    77443    77977: contig of 535 bp in length
*    77978    78057: gap of unknown length
*    78058    78637: contig of 580 bp in length
*    78638    78717: gap of unknown length
*    78718    79315: contig of 598 bp in length
*    79316    79395: gap of unknown length
*    79396    79957: contig of 562 bp in length
*    79958    80037: gap of unknown length
*    80038    80576: contig of 539 bp in length
*    80577    80656: gap of unknown length
*    80657    81230: contig of 574 bp in length
*    81231    81310: gap of unknown length
*    81311    81885: contig of 575 bp in length
*    81886    81965: gap of unknown length
*    81966    82431: contig of 466 bp in length
*    82432    82511: gap of unknown length
*    82512    82956: contig of 445 bp in length
*    82957    83036: gap of unknown length
*    83037    83606: contig of 570 bp in length
*    83607    83686: gap of unknown length
*    83687    84100: contig of 414 bp in length
*    84101    84180: gap of unknown length
*    84181    84745: contig of 565 bp in length
*    84746    84825: gap of unknown length
*    84826    85467: contig of 642 bp in length
*    85468    85547: gap of unknown length
*    85548    86188: contig of 641 bp in length
*    86189    86268: gap of unknown length
*    86269    86784: contig of 516 bp in length
*    86785    86864: gap of unknown length
*    86865    87341: contig of 477 bp in length
*    87342    87421: gap of unknown length
*    87422    88052: contig of 631 bp in length
*    88053    88132: gap of unknown length
*    88133    88770: contig of 638 bp in length
*    88771    88850: gap of unknown length
*    88851    89261: contig of 411 bp in length
*    89262    89341: gap of unknown length
*    89342    89943: contig of 602 bp in length
*    89944    90023: gap of unknown length
*    90024    90481: contig of 458 bp in length
*    90482    90561: gap of unknown length
*    90562    90891: contig of 330 bp in length
*    90892    90971: gap of unknown length
*    90972    91319: contig of 348 bp in length
*    91320    91399: gap of unknown length
*    91400    91859: contig of 460 bp in length
*    91860    91939: gap of unknown length
*    91940    92488: contig of 549 bp in length
*    92489    92568: gap of unknown length
*    92569    93304: contig of 736 bp in length
*    93305    93384: gap of unknown length
*    93385    93741: contig of 357 bp in length
*    93742    93821: gap of unknown length
*    93822    94283: contig of 462 bp in length
*    94284    94363: gap of unknown length
*    94364    94904: contig of 541 bp in length
*    94905    94984: gap of unknown length
*    94985    95668: contig of 684 bp in length.
FEATURES             Location/Qualifiers
     source          1. .95668
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2"
                     /map="40A-40C"
                     /clone="BACR42E05 (D1040) RPCI-98 42.E.5"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute

  Query Match             88.0%;  Score 17.6;  DB 61;  Length 95668;
  Best Local Similarity   85.0%;  Pred. No. 2.5e+02;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ttgstcaaaatttgascaaa 20
           |||:|||||||||||: |||
Db    8602 TTGCTCAAAATTTGAGAAAA 8621


RESULT  15
AL356736
LOCUS       AL356736   108399 bp    DNA             HTG       23-JAN-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-422P10, *** SEQUENCING IN
            PROGRESS ***, 26 unordered pieces.
ACCESSION   AL356736
VERSION     AL356736.3  GI:9926625
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 108399)
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 25, 2000 this sequence version replaced gi:9213961.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA422P10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 95168 bases at least Q40
            Consensus quality: 100303 bases at least Q30
            Consensus quality: 102890 bases at least Q20
            Insert size: 105899; sum-of-contigs
            Insert size: 156533; 8.5% error; agarose-fp
            Quality coverage: 2.72x in Q20 bases; sum-of-contigs Quality
            coverage: 2.16x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1     2189: contig of 2189 bp in length
            *   2190  2289: gap of      100 bp
            *   2290    10136: contig of 7847 bp in length
            *  10137  10236: gap of      100 bp
            *  10237    14836: contig of 4600 bp in length
            *  14837  14936: gap of      100 bp
            *  14937    20068: contig of 5132 bp in length
            *  20069  20168: gap of      100 bp
            *  20169    23196: contig of 3028 bp in length
            *  23197  23296: gap of      100 bp
            *  23297    25747: contig of 2451 bp in length
            *  25748  25847: gap of      100 bp
            *  25848    28603: contig of 2756 bp in length
            *  28604  28703: gap of      100 bp
            *  28704    33157: contig of 4454 bp in length
            *  33158  33257: gap of      100 bp
            *  33258    38474: contig of 5217 bp in length
            *  38475  38574: gap of      100 bp
            *  38575    44658: contig of 6084 bp in length
            *  44659  44758: gap of      100 bp
            *  44759    49335: contig of 4577 bp in length
            *  49336  49435: gap of      100 bp
            *  49436    51799: contig of 2364 bp in length
            *  51800  51899: gap of      100 bp
            *  51900    54349: contig of 2450 bp in length
            *  54350  54449: gap of      100 bp
            *  54450    60488: contig of 6039 bp in length
            *  60489  60588: gap of      100 bp
            *  60589    67658: contig of 7070 bp in length
            *  67659  67758: gap of      100 bp
            *  67759    69787: contig of 2029 bp in length
            *  69788  69887: gap of      100 bp
            *  69888    77142: contig of 7255 bp in length
            *  77143  77242: gap of      100 bp
            *  77243    81918: contig of 4676 bp in length
            *  81919  82018: gap of      100 bp
            *  82019    86773: contig of 4755 bp in length
            *  86774  86873: gap of      100 bp
            *  86874    88965: contig of 2092 bp in length
            *  88966  89065: gap of      100 bp
            *  89066    91615: contig of 2550 bp in length
            *  91616  91715: gap of      100 bp
            *  91716    95214: contig of 3499 bp in length
            *  95215  95314: gap of      100 bp
            *  95315    99815: contig of 4501 bp in length
            *  99816  99915: gap of      100 bp
            *  99916   101937: contig of 2022 bp in length
            * 101938 102037: gap of      100 bp
            * 102038   105570: contig of 3533 bp in length
            * 105571 105670: gap of      100 bp
            * 105671   108399: contig of 2729 bp in length.
FEATURES             Location/Qualifiers
     source          1. .108399
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-422P10"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .2189
                     /note="assembly_fragment:00802
                     fragment_chain:1"
     misc_feature    2290. .10136
                     /note="assembly_fragment:01048
                     fragment_chain:1"
     misc_feature    10237. .14836
                     /note="assembly_fragment:00280
                     fragment_chain:1"
     misc_feature    14937. .20068
                     /note="assembly_fragment:00172
                     fragment_chain:1"
     misc_feature    20169. .23196
                     /note="assembly_fragment:00258
                     fragment_chain:1"
     misc_feature    23297. .25747
                     /note="assembly_fragment:00737
                     fragment_chain:1"
     misc_feature    25848. .28603
                     /note="assembly_fragment:00087
                     fragment_chain:2"
     misc_feature    28704. .33157
                     /note="assembly_fragment:00528
                     fragment_chain:2"
     misc_feature    33258. .38474
                     /note="assembly_fragment:00740
                     fragment_chain:2"
     misc_feature    38575. .44658
                     /note="assembly_fragment:00210
                     fragment_chain:3"
     misc_feature    44759. .49335
                     /note="assembly_fragment:00597
                     fragment_chain:3"
     misc_feature    49436. .51799
                     /note="assembly_fragment:01053
                     fragment_chain:3"
     misc_feature    51900. .54349
                     /note="assembly_fragment:00005
                     fragment_chain:4"
     misc_feature    54450. .60488
                     /note="assembly_fragment:00750
                     fragment_chain:4"
     misc_feature    60589. .67658
                     /note="assembly_fragment:00371
                     fragment_chain:5"
     misc_feature    67759. .69787
                     /note="assembly_fragment:01118
                     fragment_chain:5"
     misc_feature    69888. .77142
                     /note="assembly_fragment:00679
                     fragment_chain:6"
```

```
     misc_feature    77243. .81918
                     /note="assembly_fragment:00766
                     fragment_chain:6"
     misc_feature    82019. .86773
                     /note="assembly_fragment:00459"
     misc_feature    86874. .88965
                     /note="assembly_fragment:00571"
     misc_feature    89066. .91615
                     /note="assembly_fragment:00655"
     misc_feature    91716. .95214
                     /note="assembly_fragment:00708"
     misc_feature    95315. .99815
                     /note="assembly_fragment:00875"
     misc_feature    99916. .101937
                     /note="assembly_fragment:00967"
     misc_feature    102038. .105570
                     /note="assembly_fragment:01044"
     misc_feature    105671. .108399
                     /note="assembly_fragment:00130
                     clone_end:T7
                     vector_side:right"
BASE COUNT    31838 a  21618 c  20968 g  31460 t   2515 others
ORIGIN


  Query Match             88.0%;  Score 17.6;  DB 80;  Length 108399;
  Best Local Similarity   85.0%;  Pred. No. 2.5e+02;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||:|||||| ||||:||||
Db  72246 TTGCTCAAAAGTTGAGCAAA 72265


RESULT  16
CEY39B6_1
WPCOMMENT
  Sequence split into 4 fragments  LOCUS CEY39B6 Accession Z95399
     Fragment Name      Begin     End
     CEY39B6_0              1     110000
     CEY39B6_1         100001     210000
     CEY39B6_2         200001     310000
     CEY39B6_3         300001     356870
Continuation (2 of 4) of CEY39B6 from base 100001 (Z95399 Caenorhabditis elegans chromos


  Query Match             88.0%;  Score 17.6;  DB 83;  Length 110000;
  Best Local Similarity   85.0%;  Pred. No. 2.5e+02;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||: ||||||||||:||||
Db  82690 TTGGACAAAATTTGACCAAA 82709


RESULT  17
HSDJ543C6/c
LOCUS       HSDJ543C6  143475 bp    DNA             HTG       11-FEB-2001
DEFINITION  Homo sapiens chromosome 1 clone RP4-543C6 map p31.2-32.1, ***
            SEQUENCING IN PROGRESS ***, 9 unordered pieces.
ACCESSION   AL109926
VERSION     AL109926.3  GI:12750965
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 143475)
  AUTHORS   Pavitt,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk

COMMENT     On Feb 12, 2001 this sequence version replaced gi:9211584.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: dJ543C6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 88% of reads
            Chemistry: Dye-primer-amersham; 11% of reads
            Consensus quality: 139941 bases at least Q40
            Consensus quality: 140762 bases at least Q30
            Consensus quality: 141485 bases at least Q20
            Insert size: 142675; sum-of-contigs
            Insert size: 126673; 5.4% error; agarose-fp
            Quality coverage: 7.44x in Q20 bases; sum-of-contigs Quality
            coverage: 8.66x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    13724: contig of 13724 bp in length
            *    13725 13824: gap of      100 bp
            *    13825    20055: contig of 6231 bp in length
            *    20056 20155: gap of      100 bp
            *    20156    26391: contig of 6236 bp in length
            *    26392 26491: gap of      100 bp
            *    26492    33378: contig of 6887 bp in length
            *    33379 33478: gap of      100 bp
            *    33479    44248: contig of 10770 bp in length
            *    44249 44348: gap of      100 bp
            *    44349    56521: contig of 12173 bp in length
            *    56522 56621: gap of      100 bp
            *    56622    58919: contig of 2298 bp in length
            *    58920 59019: gap of      100 bp
            *    59020    71651: contig of 12632 bp in length
            *    71652 71751: gap of      100 bp
            *    71752   143475: contig of 71724 bp in length.
FEATURES             Location/Qualifiers
     source          1. .143475
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p31.2-32.1"
                     /clone="RP4-543C6"
                     /clone_lib="RPCI-4"
     misc_feature    1. .13724
                     /note="assembly_fragment:01333
                     fragment_chain:1"
     misc_feature    13825. .20055
                     /note="assembly_fragment:00310
                     fragment_chain:1"
     misc_feature    20156. .26391
                     /note="assembly_fragment:02156
                     fragment_chain:2"
     misc_feature    26492. .33378
                     /note="assembly_fragment:01946
                     fragment_chain:2"
     misc_feature    33479. .44248
                     /note="assembly_fragment:00523"
     misc_feature    44349. .56521
                     /note="assembly_fragment:00764"
     misc_feature    56622. .58919
                     /note="assembly_fragment:01536"
     misc_feature    59020. .71651
```

```
                     /note="assembly_fragment:01955"
     misc_feature    71752. .143475
                     /note="assembly_fragment:02758"
BASE COUNT    43395 a  28181 c  27614 g  43482 t    803 others
ORIGIN


  Query Match             88.0%;  Score 17.6;  DB 83;  Length 143475;
  Best Local Similarity   85.0%;  Pred. No. 2.5e+02;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||:|||||| ||||:||||
Db  81358 TTGCTCAAAAGTTGAGCAAA 81339


RESULT  18
AL158081/c
LOCUS       AL158081   159425 bp    DNA             HTG       23-JAN-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-358A9 map q31.3-32.1, ***
            SEQUENCING IN PROGRESS ***, 10 unordered pieces.
ACCESSION   AL158081
VERSION     AL158081.3  GI:9796653
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159425)
  AUTHORS   Pavitt,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 12, 2000 this sequence version replaced gi:9212544.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA358A9
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 152267 bases at least Q40
            Consensus quality: 155093 bases at least Q30
            Consensus quality: 157021 bases at least Q20
            Insert size: 158525; sum-of-contigs
            Insert size: 157428; 7.1% error; agarose-fp
            Quality coverage: 4.34x in Q20 bases; sum-of-contigs Quality
            coverage: 4.50x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    39046: contig of 39046 bp in length
            *    39047 39146: gap of      100 bp
            *    39147    56828: contig of 17682 bp in length
            *    56829 56928: gap of      100 bp
            *    56929    68440: contig of 11512 bp in length
            *    68441 68540: gap of      100 bp
            *    68541    88371: contig of 19831 bp in length
            *    88372 88471: gap of      100 bp
            *    88472   115053: contig of 26582 bp in length
            *   115054 115153: gap of      100 bp
            *   115154   135644: contig of 20491 bp in length
            *   135645 135744: gap of      100 bp
            *   135745   139246: contig of 3502 bp in length
            *   139247 139346: gap of      100 bp
            *   139347   152407: contig of 13061 bp in length
            *   152408 152507: gap of      100 bp
            *   152508   154691: contig of 2184 bp in length
            *   154692 154791: gap of      100 bp
            *   154792   159425: contig of 4634 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159425
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="q31.3-32.1"
                     /clone="RP11-358A9"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .39046
                     /note="assembly_fragment:01005
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    39147. .56828
                     /note="assembly_fragment:00427
                     fragment_chain:1"
     misc_feature    56929. .68440
                     /note="assembly_fragment:00712
                     fragment_chain:1"
     misc_feature    68541. .88371
                     /note="assembly_fragment:00100
                     fragment_chain:1"
     misc_feature    88472. .115053
                     /note="assembly_fragment:00459
                     fragment_chain:2"
     misc_feature    115154. .135644
                     /note="assembly_fragment:00327
                     fragment_chain:2"
     misc_feature    135745. .139246
                     /note="assembly_fragment:01134
                     fragment_chain:3"
     misc_feature    139347. .152407
                     /note="assembly_fragment:00086
                     fragment_chain:3"
     misc_feature    152508. .154691
                     /note="assembly_fragment:00030"
     misc_feature    154792. .159425
                     /note="assembly_fragment:00905"
BASE COUNT    52091 a  28696 c  28543 g  49189 t    906 others
ORIGIN


  Query Match             88.0%;  Score 17.6;  DB 79;  Length 159425;
  Best Local Similarity   85.0%;  Pred. No. 2.4e+02;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||:|||||||| |||:||||
Db 145095 TTGGTCAAAATATGACCAAA 145076


RESULT  19
AC024697/c
LOCUS       AC024697   161641 bp    DNA             HTG       30-NOV-2000
DEFINITION  Homo sapiens chromosome 2 clone RP11-9D6, WORKING DRAFT SEQUENCE, 7
            unordered pieces.
ACCESSION   AC024697
VERSION     AC024697.4  GI:11465113
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161641)
  AUTHORS   Waterston,R.H.
```

```
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161641)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Nov 30, 2000 this sequence version replaced gi:8318595.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0009D06
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 94%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 94% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 158154 bases at least Q40
            Consensus quality: 159373 bases at least Q30
            Consensus quality: 159905 bases at least Q20
            Insert size: 150000; agarose-fp
            Insert size: 161041; sum-of-contigs
            Quality coverage: 5.45 in Q20 bases; agarose-fp
            Quality coverage: 5.26 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1685: contig of 1685 bp in length
            *    1686     1785: gap of unknown length
            *    1786     6992: contig of 5207 bp in length
            *    6993     7092: gap of unknown length
            *    7093    43524: contig of 36432 bp in length
            *   43525    43624: gap of unknown length
            *   43625    53856: contig of 10232 bp in length
            *   53857    53956: gap of unknown length
            *   53957    70038: contig of 16082 bp in length
            *   70039    70138: gap of unknown length
            *   70139   114242: contig of 44104 bp in length
            *  114243   114342: gap of unknown length
            *  114343   161641: contig of 47299 bp in length.
FEATURES             Location/Qualifiers
     source          1. .161641
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-9D6"
     misc_feature    1. .1685
                     /note="assembly_name:Contig10"
     misc_feature    1786. .6992
                     /note="assembly_name:Contig11
                     clone_end:SP6
                     vector_side:right"
     misc_feature    7093. .43524
                     /note="assembly_name:Contig14
                     clone_end:T7
                     vector_side:right"
     misc_feature    43625. .53856
                     /note="assembly_name:Contig12"
     misc_feature    53957. .70038
                     /note="assembly_name:Contig13"
     misc_feature    70139. .114242
                     /note="assembly_name:Contig15"
     misc_feature    114343. .161641
                     /note="assembly_name:Contig16"
BASE COUNT    51564 a  29165 c  29620 g  50686 t    606 others
ORIGIN


  Query Match             88.0%;  Score 17.6;  DB 69;  Length 161641;
  Best Local Similarity   85.0%;  Pred. No. 2.4e+02;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ttgstcaaaatttgascaaa 20
           |||:||||||||| :||||
Db 134782 TTGGTCAAAATTTGTCCAAA 134763


RESULT  20
AL161634
LOCUS       AL161634   163198 bp    DNA             HTG       23-JAN-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-378F16 map q31.2-32.1, ***
            SEQUENCING IN PROGRESS ***, 10 unordered pieces.
ACCESSION   AL161634
VERSION     AL161634.4  GI:9796822
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163198)
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 12, 2000 this sequence version replaced gi:9212814.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA378F16
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 160364 bases at least Q40
            Consensus quality: 161529 bases at least Q30
            Consensus quality: 161977 bases at least Q20
            Insert size: 162298; sum-of-contigs
            Insert size: 157056; 7.6% error; agarose-fp
            Quality coverage: 5.07x in Q20 bases; sum-of-contigs Quality
            coverage: 5.27x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    29583: contig of 29583 bp in length
            *   29584 29683: gap of      100 bp
            *   29684    39123: contig of 9440 bp in length
            *   39124 39223: gap of      100 bp
            *   39224    46416: contig of 7193 bp in length
            *   46417 46516: gap of      100 bp
            *   46517   110205: contig of 63689 bp in length
            *  110206 110305: gap of      100 bp
            *  110306   113056: contig of 2751 bp in length
            *  113057 113156: gap of      100 bp
            *  113157   116361: contig of 3205 bp in length
            *  116362 116461: gap of      100 bp
```

```
           *   116462   120047: contig of 3586 bp in length
           *   120048 120147: gap of      100 bp
           *   120148   133234: contig of 13087 bp in length
           *   133235 133334: gap of      100 bp
           *   133335   158006: contig of 24672 bp in length
           *   158007 158106: gap of      100 bp
           *   158107   163198: contig of 5092 bp in length.
FEATURES             Location/Qualifiers
     source          1. .163198
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="q31.2-32.1"
                     /clone="RP11-378F16"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .29583
                     /note="assembly_fragment:00818
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    29684. .39123
                     /note="assembly_fragment:01111
                     fragment_chain:1"
     misc_feature    39224. .46416
                     /note="assembly_fragment:00702
                     fragment_chain:1"
     misc_feature    46517. .110205
                     /note="assembly_fragment:01237
                     fragment_chain:1"
     misc_feature    110306. .113056
                     /note="assembly_fragment:00658
                     fragment_chain:2"
     misc_feature    113157. .116361
                     /note="assembly_fragment:00746
                     fragment_chain:2"
     misc_feature    116462. .120047
                     /note="assembly_fragment:00472
                     fragment_chain:2"
     misc_feature    120148. .133234
                     /note="assembly_fragment:00325
                     fragment_chain:2"
     misc_feature    133335. .158006
                     /note="assembly_fragment:01410
                     fragment_chain:2"
     misc_feature    158107. .163198
                     /note="assembly_fragment:00688
                     clone_end:T7
                     vector_side:right"
BASE COUNT    51149 a  29261 c  29288 g  52596 t    904 others
ORIGIN


  Query Match             88.0%;  Score 17.6;  DB 79;  Length 163198;
  Best Local Similarity   85.0%;  Pred. No. 2.4e+02;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ttgstcaaaatttgascaaa 20
           |||:|||||||  |||:||||
Db 146164 TTGGTCAAAATATGACCAAA 146183



Search completed: June  7, 2001, 18:04:28
Job time: 6261 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 7, 2001, 17:12:22 ; Search time 200.8 Seconds
(without alignments)
58.146 Million cell updates/sec

Title: US-09-601-561-1
Perfect score: 20
Sequence: 1 ttgstcaaaatttgascaaa 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : N_Geneseq_0401:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	17.6	88.0	5342	21	D00374	Rat smooth muscle-
c	2	16.6	83.0	500	21	C94123	Cat flea head and
	3	16	80.0	461	21	C56291	Pinus radiata tran
c	4	16	80.0	1351	21	F13173	Aspergillus oryzae
c	5	16	80.0	2720	20	Z33573	Human breast tumou
c	6	16	80.0	2975	20	X03792	Human semaphorin E
c	7	16	80.0	2975	21	C66800	Clone BR533_4 codi
c	8	16	80.0	5355	20	X89220	Seq ID No: 28 of W
	9	16	80.0	5355	20	X89221	Seq ID No: 30 of W
c	10	16	80.0	5355	21	Z39404	Human Trad protein
	11	16	80.0	5355	21	Z39405	Human Trad protein
	12	16	80.0	9359	20	X20262	Borrelia burgdorfe
	13	16	80.0	94618	21	F22285	BAC containing rep
	14	16	80.0	121001	21	F22284	BAC containing rep
	15	15.6	78.0	1226	21	C60027	Human secreted pro
	16	15.6	78.0	1345	21	C93450	Human secreted pro
	17	15.6	78.0	1945	20	X99684	Mouse Tcf-3 (mTcf-
	18	15.6	78.0	1082138	21	F22305	Arabidopsis thalia
c	19	15.4	77.0	829	20	Z29907	cDNA encoding a SC
	20	15.4	77.0	2720	20	Z33573	Human breast tumou
	21	15.4	77.0	2975	20	X03792	Human semaphorin E
	22	15.4	77.0	2975	21	C66800	Clone BR533_4 codi
	23	15.4	77.0	20284	21	C69139	Human ABC1 gene ex
	24	15.2	76.0	701	14	Q53862	Human Glia Maturat
	25	15.2	76.0	1121	21	C36705	Arabidopsis thalia
	26	15	75.0	466	20	X13880	Enterococcus faeca
	27	15	75.0	500	21	C94123	Cat flea head and
c	28	15	75.0	578	21	A16434	Human colon cancer
c	29	15	75.0	745	20	Z15041	Human gene express
c	30	15	75.0	745	21	A01576	Human colon cancer
c	31	15	75.0	1062	21	C48389	Arabidopsis thalia
c	32	15	75.0	1183	21	C43780	Zea mays DNA fragm
c	33	15	75.0	1207	21	C46941	Arabidopsis thalia
c	34	15	75.0	1378	21	C35123	Arabidopsis thalia
c	35	15	75.0	1392	21	C59260	Human secreted pro
c	36	15	75.0	1440	20	X61662	B. burgdorferi ant
c	37	15	75.0	1515	20	X61661	B. burgdorferi ant
	38	15	75.0	1524	21	A70107	Plasmodium falcipa
c	39	15	75.0	1642	21	C46009	Arabidopsis thalia
c	40	15	75.0	1661	21	C44031	Zea mays DNA fragm
	41	15	75.0	2040	19	V63088	Human Tcf-4B cDNA.
	42	15	75.0	2040	21	Z51337	Human Tcf-4E gene.
	43	15	75.0	2171	20	X99685	Mouse Tcf-4 (mTcf-
c	44	15	75.0	2387	19	V52189	Streptococcus pneu
	45	15	75.0	2444	19	V63089	Human Tcf-4E cDNA.
	46	15	75.0	2444	21	Z51338	Human Tcf-4B gene.
c	47	15	75.0	5342	21	D00374	Rat smooth muscle-
	48	15	75.0	7434	20	X12971	Enterococcus faeca
	49	15	75.0	111309	20	X20250	Borrelia burgdorfe
c	50	15	75.0	116277	20	X20249	Borrelia burgdorfe

ALIGNMENTS

RESULT 1
D00374
ID D00374 standard; DNA; 5342 BP.
XX
AC D00374;
XX
DT 09-AUG-2000 (first entry)
XX
DE Rat smooth muscle-alpha actin (SM alphaA) gene regulatory sequence.
XX
KW Smooth muscle alpha actin gene; SM alphaA; rat; smooth muscle cell; SMC;
KW anticancer; antiproliferative; cardiant; cis regulatory element; cancer;
KW treatment; proliferative disorder; coronary artery disease; stroke;
KW hypertension; asthma; multiple gastrointestinal disorder; gene therapy;
KW urogenital disorder; reproductive disorder; CArG; ds.
XX
OS Rattus sp.
XX
FH Key Location/Qualifiers
FT promoter 1..2558
FT /*tag= a
FT enhancer 2438..2447
FT /*tag= b
FT /label= CArg_B
FT /note= "Highly conserved cis acting element essential to
FT direct SMC-specific transcription in vivo"
FT enhancer 2488..2495
FT /*tag= c
FT /label= CArg_A

```
FT                   /note= "Highly conserved cis acting element essential to
FT                   direct SMC-specific transcription in vivo"
FT   exon            2559..3330
FT                   /*tag= d
FT                   /number= 1
FT   intron          3331..3656
FT                   /*tag= e
FT                   /number= 1
FT   enhancer        3559..3568
FT                   /*tag= f
FT                   /label= Intronic_CArg
FT                   /note= "Highly conserved cis acting element essential to
FT                   direct SMC-specific transcription in vivo"
XX
PN   WO200024254-A1.
XX
PD   04-MAY-2000.
XX
PF   23-OCT-1999;   99WO-US24972.
XX
PR   23-OCT-1998;   98US-0105330.
XX
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
XX
PI   Owens GK,  Mack C,  Blank R;
XX
DR   WPI; 2000-350519/30.
XX
PT   Isolated polynucleotide is used for inhibiting or treating smooth
PT   muscle cell-related cancer or other proliferative disorders and for
PT   preventing or delaying heart attack -
XX
PS   Claim 1a; Page 84-85; 94pp; English.
XX
CC   The present DNA sequence is the rat smooth muscle alpha actin (SM alphaA)
CC   gene 5' regulatory sequence, comprising the promoter and first intron.
CC   It corresponds to bases -2558 to +2784 of the SM alphaA gene. An intronic
CC   fragment from +773 to +1098 bp relative to the start of transcription is
CC   highly conserved and contains cis-acting elements required to direct
CC   SMC-specific transcription in vivo. This sequence has anticancer,
CC   antiproliferative and cardiant activity. The SM alphaA regulatory
CC   sequence is useful to inhibit and treat smooth muscle cell (SMC)-related
CC   cancers or other proliferative disorders. It is also used for the
CC   treatment of SMC-related disorders, like coronary artery disease, stroke,
CC   hypertension, asthma and multiple gastrointestinal, urogenital and
CC   reproductive disorders. It is also used in gene therapy.
XX
SQ   Sequence 5342 BP; 1533 A; 1178 C; 1127 G; 1504 T; 0 other;
```

```
  Query Match             88.0%;  Score 17.6;  DB 21;  Length 5342;
  Best Local Similarity   85.0%;  Pred. No. 17;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ttgstcaaaatttgascaaa 20
           |||:|||||| ||||:||||
Db     897 ttgctcaaaaattgagcaaa 916
```

```
RESULT   2
C94123/c
ID   C94123 standard; cDNA; 500 BP.
XX
AC   C94123;
XX
DT   19-FEB-2001  (first entry)
XX
DE   Cat flea head and nerve cord (HNC) cDNA, SEQ ID NO:618.
XX
KW   Cat flea; head and nerve cord nucleic acid; HNC; flea infestation;
KW   vaccine; antiparasitic; therapeutic target; diagnosis; detection; ss.
XX
OS   Ctenocephalides felis.
XX
PN   WO200061621-A2.
XX
PD   19-OCT-2000.
XX
PF   07-APR-2000; 2000WO-US09437.
XX
PR   09-APR-1999;   99US-0128704.
XX
PA   (HESK-) HESKA CORP.
XX
PI   Brandt KS,  Gaines PJ,  Stinchcomb DT,  Wisnewski N;
XX
DR   WPI; 2000-656323/63.
XX
PT   Flea Malpighian tubule and head and nerve cord tissue derived nucleic
PT   acids useful for the prevention, diagnosis and treatment of flea
PT   infestations -
XX
PS   Claim 26; Page 448; 964pp; English.
XX
CC   The invention relates to novel cat flea (Ctenocephalides felis) nucleic
CC   acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
CC   or head and nerve cord (HNC) tissue. The invention also relates to the
CC   encoded proteins. The invention additionally encompasses expression
CC   constructs, recombinant viruses and recombinant cells comprising the
CC   nucleic acids of the invention, recombinant production of the proteins,
CC   antibodies against the proteins, a method of identifying inhibitors of
CC   the proteins, and compositions comprising the inhibitors for
CC   administration to an animal. The nucleic acids, and the proteins they
CC   encode may be used in the prevention, treatment and diagnosis of diseases
CC   associated with flea infestations. For example, the nucleic acids may be
CC   used to produce an HMT or HNC protein according to standard recombinant
CC   DNA methodology by inserting the nucleic acids into a host cell and
CC   culturing the cell to express the protein. The HMT and HNC nucleic acids
CC   may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
CC   and quantitate the presence of cat flea or other homologous nucleic acid
CC   sequences in samples. They may also be used to study the expression and
CC   function of the proteins and their role in metabolism. The HMT and HNC
CC   proteins may be used as antigens in the production of specific
CC   antibodies, and in assays to identify modulators (agonists and
CC   antagonists) of HMT and/or HNC protein expression and activity. The
CC   anti-HMT/HNC protein antibodies and antagonists may also be used to
CC   downregulate protein expression and activity. The antibodies may also be
CC   used as diagnostic agents for detecting the presence of flea polypeptides
CC   in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
CC   present sequence represents a cat flea HNC cDNA of the invention.
XX
SQ   Sequence 500 BP; 165 A; 85 C; 79 G; 171 T; 0 other;
```

```
  Query Match             83.0%;  Score 16.6;  DB 21;  Length 500;
  Best Local Similarity   84.2%;  Pred. No. 42;
  Matches   16;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ttgstcaaaatttgascaa 19
           |||:|||| ||||||:|||
Db     418 TTGGTCAAGATTTGACCAA 400
```

```
RESULT   3
C56291
ID   C56291 standard; DNA; 461 BP.
XX
AC   C56291;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Pinus radiata transcription factor DNA sequence #91.
XX
KW   Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW   poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW   basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
```

KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS Pinus radiata.
XX
PN WO200053724-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06112.
XX
PR 11-MAR-1999; 99US-0266513.
PR 18-AUG-1999; 99US-0149485.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Wood M, McGrath A, Shenk MA, Glenn M;
XX
DR WPI; 2000-579369/54.
XX
PT New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide
PT -
XX
PS Claim 1; Page 154; 747pp; English.
XX
CC The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper,
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC and MYB.
XX
SQ Sequence 461 BP; 142 A; 86 C; 110 G; 123 T; 0 other;

Query Match 80.0%; Score 16; DB 21; Length 461;
Best Local Similarity 80.0%; Pred. No. 80;
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ttgstcaaaatttgascaaa 20
|| :|||||||||||: |||
Db 435 ttagtcaaaatttgagaaaa 454

RESULT 4
F13173/c
ID F13173 standard; cDNA; 1351 BP.
XX
AC F13173;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus oryzae EST SEQ ID NO:5696.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus oryzae.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 88; Page 2361; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway engineering.
CC Using ESTs provides several advantages over genomic or random cDNA
CC clones including elimination of redundancy as one spot on an array
CC equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. F07478 to F11247 represents ESTs from Fusarium
CC venenatum; F11248 to F11853 represents ESTs from Aspergillus niger;
CC F11854 to F14878 represents ESTs from Aspergillus oryzae; and F14879 to
CC F15337 represents ESTs from Trichoderma reesei, which are all
CC specifically claimed in the present invention.
XX
SQ Sequence 1351 BP; 305 A; 365 C; 315 G; 365 T; 1 other;

Query Match 80.0%; Score 16; DB 21; Length 1351;
Best Local Similarity 80.0%; Pred. No. 88;
Matches 16; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ttgstcaaaatttgascaaa 20
|||:|||| | ||||:||||
Db 1215 TTGGTCAAGACTTGACCAAA 1196

RESULT 5
Z33573/c
ID Z33573 standard; cDNA; 2720 BP.
XX
AC Z33573;
XX
DT 08-DEC-1999 (first entry)
XX
DE Human breast tumour-associated EST 33.
XX
KW Expressed sequence tag; EST; human; breast; cancer; cytostatic;
KW medicaments; gene therapy; treatment; fat metabolism; ss.
XX
OS Homo sapiens.
XX
PN DE19813835-A1.
XX

```
PD   23-SEP-1999.
XX
PF   20-MAR-1998;   98DE-1013835.
XX
PR   20-MAR-1998;   98DE-1013835.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E,  Rosenthal A;
XX
DR   WPI; 1999-528979/45.
XX
PT   Human nucleic acid sequences and protein products from normal breast
PT   tissue, useful for breast cancer therapy -
XX
PS   Claim 1a; 121; 206pp; German.
XX
CC   This invention describes novel human nucleic acid sequences from normal
CC   breast tissue which have cytostatic activity. The nucleic acid sequences
CC   can be used to produce and isolate full-length gene sequences. They can
CC   be used to express proteins, which can be used as tools to find an
CC   activity against breast cancer. The sequences can be used in sense or
CC   antisense form. They are especially useful for medicaments for gene
CC   therapy to treat breast cancer and for treating illnesses associated
CC   with fat metabolism. Z33541-Z33610 represent expressed sequence tags
CC   described in the method of the invention.
XX
SQ   Sequence 2720 BP; 947 A; 398 C; 454 G; 921 T; 0 other;


  Query Match             80.0%;  Score 16;  DB 20;  Length 2720;
  Best Local Similarity   88.9%;  Pred. No. 93;
  Matches   16;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgasca 18
          |||:||||||||||| ||
Db    388 TTGCTCAAAATTTGATCA 371


RESULT   6
X03792/c
ID   X03792 standard; cDNA; 2975 BP.
XX
AC   X03792;
XX
DT   31-MAR-1999  (first entry)
XX
DE   Human semaphorin E encoding cDNA clone BR5334.
XX
KW   Human; semaphorin E; clone BR5334; nutritional; immune stimulating;
KW   vaccine; haematopoiesis regulating activity; tissue growth;
KW   haemostatic; thrombolytic; anti-inflammatory; tumour inhibition;
KW   gene therapy; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             245..2500
FT                   /*tag=  a
XX
PN   WO9853065-A1.
XX
PD   26-NOV-1998.
XX
PF   19-MAY-1998;   98WO-US10188.
XX
PR   18-MAY-1998;   98US-0080695.
PR   19-MAY-1997;   97US-0858834.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Fechtel K,  Howes SH,  Jacobs K,  Lavallie ER;
PI   McCoy JM,  Merberg D,  Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-059742/05.
DR   P-PSDB; W30617.
XX
PT   New polypeptides encoding secreted human proteins - derived from a
PT   human foetal kidney cDNA library
XX
PS   Claim 1; Page 43-44; 58pp; English.
XX
CC   The present sequence encodes human semaphorin E from cDNA clone BR5334.
CC   Human semaphorin E polynucleotide sequences and protein sequences from
CC   the present invention, are predicted to have biological activities which
CC   would make them suitable for treating, preventing or ameliorating
CC   medical conditions in humans and animals, although no supporting data is
CC   given. Suggested activities include nutritional activity, immune
CC   stimulating (e.g. as vaccines) or suppressing activity, haematopoiesis
CC   regulating activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, and tumour inhibition
CC   activity. The polynucleotide sequences are also stated to be useful for
CC   gene therapy. A host cell transfected with the polynucleotide sequence
CC   encoding human semaphorin E or its subfragments and variants is useful
CC   for recombinant production of the clone BR5334 related protein.
XX
SQ   Sequence 2975 BP; 887 A; 620 C; 665 G; 802 T; 1 other;


  Query Match             80.0%;  Score 16;  DB 20;  Length 2975;
  Best Local Similarity   88.9%;  Pred. No. 94;
  Matches   16;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgasca 18
          |||:||||||||||| ||
Db   2569 TTGCTCAAAATTTGATCA 2552


RESULT   7
C66800/c
ID   C66800 standard; cDNA; 2975 BP.
XX
AC   C66800;
XX
DT   19-FEB-2001  (first entry)
XX
DE   Clone BR533_4 coding sequence.
XX
KW   Human; clone BR533_4; secreted protein; immune deficiency;
KW   microbial infection; autoimmune disorder; allergy; asthma; inflammation;
KW   myeloid deficiency; lymphoid cell deficiency; anaemia; burn;
KW   wound healing; ulcer; periodontal disease; nervous system disease;
KW   neuropathy; lung fibrosis; liver fibrosis; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200063692-A1.
XX
PD   26-OCT-2000.
XX
PF   14-APR-2000; 2000WO-US10048.
XX
PR   15-APR-1999;   99US-0292550.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  McCoy JM,  Lavallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M,  Spaulding V;
XX
DR   WPI; 2000-679620/66.
DR   P-PSDB; B28379.
XX
PT   New monoclonal antibodies, useful for treating cancer and
PT   immunodetection of secreted proteins which are in turn useful for
```

```
PT   treating neurological, inflammatory, immune diseases and microbial
PT   infections -
XX
PS   Disclosure; Pages 66-67; 75pp; English.
XX
CC   The present sequence is the coding sequence for human clone BR533_4. The
CC   present sequence was isolated from a human foetal kidney cDNA library.
CC   The BR533_4 protein is a secreted protein and can be used to treat a
CC   number of conditions including various immune deficiencies and disorders,
CC   microbial infections, autoimmune disorders, allergic reactions such as
CC   asthma, respiratory problems, inflammation, myeloid or lymphoid cell
CC   deficiencies, anaemias, burns, wound healing, ulcers, periodontal
CC   disease, central and peripheral nervous system diseases and neuropathies,
CC   lung or liver fibrosis and cancer.
XX
SQ   Sequence 2975 BP; 887 A; 620 C; 665 G; 802 T; 1 other;

  Query Match             80.0%;  Score 16;  DB 21;  Length 2975;
  Best Local Similarity   88.9%;  Pred. No. 94;
  Matches   16;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ttgstcaaaatttgasca 18
           |||:||||||||||| ||
Db    2569 TTGCTCAAAATTTGATCA 2552


RESULT   8
X89220/c
ID   X89220 standard; DNA; 5355 BP.
XX
AC   X89220;
XX
DT   15-SEP-1999  (first entry)
XX
DE   Seq ID No: 28 of WO9933961.
XX
KW   DRAK1; DRAK2; DAP kinase related apoptosis inducing kinase; human;
KW   apoptosis; breast cancer; ovarian cancer; lymphoma; autoimmune disease;
KW   viral infection; adenovirus; poxvirus; HIV; Alzheimer's disease;
KW   Parkinson's disease; arteriosclerosis; alcoholism; rheumatoid arthritis;
KW   diabetes; ss.
XX
OS   Homo sapiens.
XX
PN   WO9933961-A1.
XX
PD   08-JUL-1999.
XX
PF   25-DEC-1998;   98WO-JP05974.
XX
PR   17-APR-1998;   98JP-0108150.
PR   26-DEC-1997;   97JP-0367640.
PR   26-DEC-1997;   97JP-0367641.
PR   17-APR-1998;   98JP-0108149.
XX
PA   (ASAH ) ASAHI KASEI KOGYO KK.
XX
PI   Akira S,  Kawai T;
XX
DR   WPI; 1999-430239/36.
DR   P-PSDB; Y27163.
XX
PT   New kinase with apoptosis induction activity useful in the treatment
PT   of cancer, autoimmune diseases and viral infections
XX
PS   Disclosure; Page 157-166; 180pp; Japanese.
XX
CC   The invention provides kinases DRAK1 and DRAK2 (DAP kinase related
CC   apoptosis inducing kinase) having apoptosis inducing activity. The
CC   kinases can be expressed recombinantly by transforming host cells with
CC   vectors comprising the nucleic acids encoding the kinases. The kinases
CC   are useful in the treatment, prevention, diagnosis and investigation of
CC   diseases with which apoptosis is associated, such as hormonally regulated
CC   cancer (such as breast cancer, ovarian cancer, lymphoma); autoimmune
CC   diseases; viral infections (such as herpes, adenovirus, poxvirus, HIV);
CC   Alzheimer's disease; Parkinson's disease; arteriosclerosis; alcoholism;
CC   rheumatoid arthritis; and diabetes.
XX
SQ   Sequence 5355 BP; 1515 A; 1278 C; 1302 G; 1260 T; 0 other;

  Query Match             80.0%;  Score 16;  DB 20;  Length 5355;
  Best Local Similarity   80.0%;  Pred. No. 99;
  Matches   16;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ttgstcaaaatttgascaaa 20
           |||:| ||||||| |:||||
Db    4369 TTGCTTAAAATTTCAGCAAA 4350


RESULT   9
X89221
ID   X89221 standard; DNA; 5355 BP.
XX
AC   X89221;
XX
DT   15-SEP-1999  (first entry)
XX
DE   Seq ID No: 30 of WO9933961.
XX
KW   DRAK1; DRAK2; DAP kinase related apoptosis inducing kinase; human;
KW   apoptosis; breast cancer; ovarian cancer; lymphoma; autoimmune disease;
KW   viral infection; adenovirus; poxvirus; HIV; Alzheimer's disease;
KW   Parkinson's disease; arteriosclerosis; alcoholism; rheumatoid arthritis;
KW   diabetes; ss.
XX
OS   Homo sapiens.
XX
PN   WO9933961-A1.
XX
PD   08-JUL-1999.
XX
PF   25-DEC-1998;   98WO-JP05974.
XX
PR   17-APR-1998;   98JP-0108150.
PR   26-DEC-1997;   97JP-0367640.
PR   26-DEC-1997;   97JP-0367641.
PR   17-APR-1998;   98JP-0108149.
XX
PA   (ASAH ) ASAHI KASEI KOGYO KK.
XX
PI   Akira S,  Kawai T;
XX
DR   WPI; 1999-430239/36.
XX
PT   New kinase with apoptosis induction activity useful in the treatment
PT   of cancer, autoimmune diseases and viral infections
XX
PS   Disclosure; Page 173-176; 180pp; Japanese.
XX
CC   The invention provides kinases DRAK1 and DRAK2 (DAP kinase related
CC   apoptosis inducing kinase) having apoptosis inducing activity. The
CC   kinases can be expressed recombinantly by transforming host cells with
CC   vectors comprising the nucleic acids encoding the kinases. The kinases
CC   are useful in the treatment, prevention, diagnosis and investigation of
CC   diseases with which apoptosis is associated, such as hormonally regulated
CC   cancer (such as breast cancer, ovarian cancer, lymphoma); autoimmune
CC   diseases; viral infections (such as herpes, adenovirus, poxvirus, HIV);
CC   Alzheimer's disease; Parkinson's disease; arteriosclerosis; alcoholism;
CC   rheumatoid arthritis; and diabetes.
XX
SQ   Sequence 5355 BP; 1260 A; 1302 C; 1278 G; 1515 T; 0 other;

  Query Match             80.0%;  Score 16;  DB 20;  Length 5355;
```

```
  Best Local Similarity   80.0%;  Pred. No. 99;
  Matches   16;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||:| ||||||| |:||||
Db    987 ttgcttaaaatttcagcaaa 1006


RESULT  10
Z39404/c
ID   Z39404 standard; cDNA to mRNA; 5355 BP.
XX
AC   Z39404;
XX
DT   23-FEB-2000  (first entry)
XX
DE   Human Trad protein encoding nucleotide sequence.
XX
KW   Trad protein; phosphoenzymatic activity; diagnosis; Rho subfamily;
KW   skeletal muscle; human; ss.
XX
OS   Homo sapiens.
XX
PN   JP11290083-A.
XX
PD   26-OCT-1999.
XX
PF   17-APR-1998;   98JP-0108151.
XX
PR   17-APR-1998;   98JP-0108151.
XX
PA   (ASAH ) ASAHI KASEI KOGYO KK.
XX
DR   WPI; 2000-016982/02.
DR   P-PSDB; Y56781.
XX
PT   A new protein Trad or its salts - useful for the screening of diseases
PT   involving the Rho subfamily on the function of skeletal muscles
XX
PS   Claim 4; Page 13-18; 25pp; Japanese.
XX
CC   This DNA encodes a human protein Trad. The protein is useful for the
CC   screening and the evaluation of an inducer or an inhibitor of
CC   phosphoenzymatic activity of Trad and for the diagnosis of diseases
CC   participated by Rho subfamily on the function of skeletal muscles.
XX
SQ   Sequence 5355 BP; 1515 A; 1278 C; 1302 G; 1260 T; 0 other;


  Query Match             80.0%;  Score 16;  DB 21;  Length 5355;
  Best Local Similarity   80.0%;  Pred. No. 99;
  Matches   16;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||:| ||||||| |:||||
Db   4369 TTGCTTAAAATTTCAGCAAA 4350


RESULT  11
Z39405
ID   Z39405 standard; cDNA to mRNA; 5355 BP.
XX
AC   Z39405;
XX
DT   23-FEB-2000  (first entry)
XX
DE   Human Trad protein DNA complementary nucleotide sequence.
XX
KW   Trad protein; phosphoenzymatic activity; diagnosis; Rho subfamily;
KW   skeletal muscle; human; complementary; ss.
XX
OS   Homo sapiens.
XX
PN   JP11290083-A.
XX
PD   26-OCT-1999.
XX
PF   17-APR-1998;   98JP-0108151.
XX
PR   17-APR-1998;   98JP-0108151.
XX
PA   (ASAH ) ASAHI KASEI KOGYO KK.
XX
DR   WPI; 2000-016982/02.
XX
PT   A new protein Trad or its salts - useful for the screening of diseases
PT   involving the Rho subfamily on the function of skeletal muscles
XX
PS   Claim 9; Page 19-20; 25pp; Japanese.
XX
CC   The invention provides a human protein Trad. The protein is useful for
CC   the screening and the evaluation of an inducer or an inhibitor of
CC   phosphoenzymatic activity of Trad and for the diagnosis of diseases
CC   participated by Rho subfamily on the function of skeletal muscles. The
CC   present sequence represents the complementary sequence of the Trad DNA.
XX
SQ   Sequence 5355 BP; 1260 A; 1302 C; 1278 G; 1515 T; 0 other;


  Query Match             80.0%;  Score 16;  DB 21;  Length 5355;
  Best Local Similarity   80.0%;  Pred. No. 99;
  Matches   16;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||:| ||||||| |:||||
Db    987 ttgcttaaaatttcagcaaa 1006


RESULT  12
X20262
ID   X20262 standard; DNA; 9359 BP.
XX
AC   X20262;
XX
DT   04-MAY-1999  (first entry)
XX
DE   Borrelia burgdorferi polynucleotide sequence #15.
XX
KW   Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW   epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW   infection; diagnosis; characterisation; detection; ds.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9858943-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998;   98WO-US12764.
XX
PR   03-SEP-1997;   97US-0057483.
PR   20-JUN-1997;   97US-0050359.
PR   22-JUL-1997;   97US-0053344.
PR   22-JUL-1997;   97US-0053377.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Clayton R,  Dougherty BA,  Fraser C,  Lathigra R,  Smith HO;
PI   White OR;
XX
DR   WPI; 1999-081217/07.
XX
PT   New isolated Borrelia burgdorferi nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of infections, particularly Lyme disease
```

```
XX
PS   Claim 1; Page 931-936; 1128pp; English.
XX
CC   X20248 to X20402 represent polynucleotide sequences isolated from
CC   Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC   the detection, diagnosis, characterisation, prevention and therapy of
CC   Bb infections, e.g. Lyme disease. They can also be used for the
CC   production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC   to a family of motile, spiral-shaped bacteria called Spirochetes.
CC   Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC   endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC   Lyme disease.
XX
SQ   Sequence 9359 BP; 2959 A; 1311 C; 1204 G; 3884 T; 1 other;


  Query Match             80.0%;  Score 16;  DB 20;  Length 9359;
  Best Local Similarity   80.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          || :|||||||||||: |||
Db   3612 tttctcaaaatttgacaaaa 3631


RESULT  13
F22285
ID   F22285 standard; DNA; 94618 BP.
XX
AC   F22285;
XX
DT   20-MAR-2001  (first entry)
XX
DE   BAC containing repeats from centromeres 1-4 #8.
XX
KW   Centromere; michrosome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200055325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US07392.
XX
PR   18-MAR-1999;   99US-0125219.
PR   01-APR-1999;   99US-0127409.
PR   18-MAY-1999;   99US-0134770.
PR   13-SEP-1999;   99US-0153584.
PR   17-SEP-1999;   99US-0154603.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D,  Copenhaver G,  Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited michrosomes which can serve as vectors for
PT   the construction of transgenic plant and animal cells  -
XX
PS   Claim 102; Page 431-453; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful for
CC   producing stably inherited michrosomes which can serve as vectors for
CC   the construction of transgenic plant and animal cells expressing
CC   selected proteins such as hormones, enzymes, interleukins, clotting
CC   factors, cytokines, antibodies, and growth factors.
XX
SQ   Sequence 94618 BP; 29114 A; 17215 C; 17486 G; 30801 T; 2 other;

  Query Match             80.0%;  Score 16;  DB 21;  Length 94618;
  Best Local Similarity   80.0%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||:|  ||||||||:||||
Db  87904 ttggttcaaatttgaccaaa 87923


RESULT  14
F22284
ID   F22284 standard; DNA; 121001 BP.
XX
AC   F22284;
XX
DT   20-MAR-2001  (first entry)
XX
DE   BAC containing repeats from centromeres 1-4 #7.
XX
KW   Centromere; michrosome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200055325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US07392.
XX
PR   18-MAR-1999;   99US-0125219.
PR   01-APR-1999;   99US-0127409.
PR   18-MAY-1999;   99US-0134770.
PR   13-SEP-1999;   99US-0153584.
PR   17-SEP-1999;   99US-0154603.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D,  Copenhaver G,  Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited michrosomes which can serve as vectors for
PT   the construction of transgenic plant and animal cells  -
XX
PS   Claim 102; Page 404-431; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful for
CC   producing stably inherited michrosomes which can serve as vectors for
CC   the construction of transgenic plant and animal cells expressing
CC   selected proteins such as hormones, enzymes, interleukins, clotting
CC   factors, cytokines, antibodies, and growth factors.
XX
SQ   Sequence 121001 BP; 39743 A; 24748 C; 22105 G; 33637 T; 768 other;


  Query Match             80.0%;  Score 16;  DB 21;  Length 121001;
  Best Local Similarity   80.0%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          | |:|||||||||||:|| |
Db  16250 taggtcaaaatttgagcaga 16269


RESULT  15
C60027
ID   C60027 standard; cDNA; 1226 BP.
XX
AC   C60027;
XX
DT   26-JAN-2001  (first entry)
```

XX
DE Human secreted protein gene 3 SEQ ID NO:13.
XX
KW Human; secreted protein; neuroprotective; cytostatic; cardioactive;
KW immunomodulatory; muscular; vulnerary; gastrointestinal; nephrotropic;
KW antiinfective; gynaecological; antibacterial; neural disorder; cancer;
KW immune disease; reproductive disorder; proliferative disorder;
KW gastrointestinal disease; wound healing; infectious disease;
KW food additive; chromosome 3; ss.
XX
OS Homo sapiens.
XX
PN WO200056766-A1.
XX
PD 28-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-US06824.
XX
PR 19-MAR-1999; 99US-0125359.
PR 03-DEC-1999; 99US-0168664.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-594574/56.
DR P-PSDB; B34856.
XX
PT Human secreted proteins and gene sequences encoding them, useful for
PT detection, prevention, and treatment of various disorders such as
PT cancer and immune system disorders -
XX
PS Claim 1; Page 351; 442pp; English.
XX
CC The polynucleotide sequences given in C60025-C60071 encode the human
CC secreted proteins represented in B34854-B34900. Sequences B34901-B34976
CC are fragments of proteins encoded by the genes, and also proteins with
CC which they share sequence homology. The proteins have activities based on
CC the tissues in which their encoding genes are expressed. Examples of
CC the proteins activities include: neuroprotective; cytostatic;
CC cardioactive; immunomodulatory; general muscular activity; vulnerary;
CC general gastrointestinal activity; nephrotropic; antiinfective;
CC gynaecological; and antibacterial. The human secreted proteins,
CC polynucleotides, antagonists and antagonists of the invention may be
CC useful in treating, preventing and/or diagnosing various diseases,
CC disorders and conditions such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal and proliferative
CC disorders and cancer. They may also be used in the treatment of wounds,
CC and infectious diseases. The polypeptides may be used as a food additive
CC or preservative to increase storage capabilities. Sequences C60016-C60024
CC and B34853 are used in the course of the invention during the
CC identification and characterisation of the protein and nucleotide
CC sequences.
XX
SQ Sequence 1226 BP; 462 A; 266 C; 237 G; 257 T; 4 other;

Query Match 78.0%; Score 15.6; DB 21; Length 1226;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 gstcaaaatttgascaaa 20
|:|||| ||||||:||||
Db 1058 gctcaacatttgaccaaa 1075

RESULT 16
C93450
ID C93450 standard; cDNA; 1345 BP.
XX
AC C93450;
XX
DT 16-FEB-2001 (first entry)

XX
DE Human secreted protein gene 29 SEQ ID NO:39.
XX
KW Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW nootropic; neuroprotective; antibacterial; virucide; fungicide;
KW opthalmalogical; vulnerary; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorders; cancer; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; nervous system disorder;
KW Alzheimer's disease; ocular disorder; wound healing; skin aging; ss.
XX
OS Homo sapiens.
XX
PN WO200061625-A1.
XX
PD 19-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US08981.
XX
PR 09-APR-1999; 99US-0128701.
PR 20-JAN-2000; 2000US-0177166.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-619226/59.
DR P-PSDB; B51758.
XX
PT New nucleic acid molecules encoding 48 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
PS Claim 1; Page 432-433; 500pp; English.
XX
CC Polynucleotide sequences C93422 - C93449 represent cDNA encoding human
CC secreted proteins B51724 - B51777. Sequences B51778 - B51825 represent
CC alternative polypeptides encoded by the genes, and amino acid sequences
CC to which they are homologous. The genes and proteins have activities
CC dependent on the tissues and cells in which they are expressed. Examples
CC of their activities include immunosuppressive; antiarthritic;
CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
CC cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
CC fungicide; opthalmalogical; and vulnerary. The secreted proteins,
CC polynucleotides, antagonists and agonists may be useful in treating,
CC preventing and/or diagnosing diseases and disorders such as autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
CC The polypeptides can also be used to aid wound healing and epithelial
CC cell proliferation, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, for supporting cell culture of primary
CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
CC also be used as a food additive or preservative to increase or decrease
CC storage capabilities, fat content, lipid, protein, carbohydrate,
CC vitamins, minerals, cofactors and other nutritional components.
CC Oligonucleotide C93413 - C93421 and peptide B51723 are used in the
CC isolation and characterisation of the proteins and polynucleotides of the
CC invention.
XX
SQ Sequence 1345 BP; 331 A; 408 C; 269 G; 326 T; 11 other;

Query Match 78.0%; Score 15.6; DB 21; Length 1345;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ttgstcaaaatttgasca 18
|||:|||| ||||||:||
Db 410 ttggtcaatatttgacca 427

```
RESULT 17
X99684
ID   X99684 standard; DNA; 1945 BP.
XX
AC   X99684;
XX
DT   05-OCT-1999  (first entry)
XX
DE   Mouse Tcf-3 (mTcf-3) protein encoding nucleotide.
XX
KW   Mammalian; transcription factor; Tcf-3; Tcf-4; beta-catenin; mouse;
KW   binding domain; Wnt signalling; Wnt/Wingless cascade; anticancer drug;
KW   transactivation; fetal cancer; testis cancer; ovary teratoma; mammary;
KW   myeloblastoma; medulloblastoma; intestinal cancer; response element; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1755
FT                   /*tag= a
FT                   /transl_except= (pos 13..15, aa:Asn)
FT                   /transl_except= (pos 673..675, aa:Pro)
FT                   /product= "mouse Tcf-3"
XX
PN   EP939122-A1.
XX
PD   01-SEP-1999.
XX
PF   27-FEB-1998;   98EP-0200621.
XX
PR   27-FEB-1998;   98EP-0200621.
XX
PA   (UYUT-) RIJKSUNIV UTRECHT.
XX
DR   WPI; 1999-471050/40.
DR   P-PSDB; Y28992.
XX
PT   New mouse transcription factors, useful for diagnosing and treating
PT   cancers and screening for anticancer drugs
XX
PS   Disclosure; Fig 9a; 56pp; English.
XX
CC   The invention relates to mammalian transcription factors Tcf-3 or Tcf-4
CC   protein or their derivatives. The Tcf proteins of the invention comprise
CC   a beta-catenin binding domain and are capable of complexing with beta-
CC   catenin in response to signaling through the Wnt/Wingless cascade. The
CC   Tcf-3 and Tcf-4 genes, proteins and antibodies against the proteins may
CC   be used to screen for candidate anticancer drugs by identifying compounds
CC   which interact with beta-catenin/Tcf-3 or beta-catenin/Tcf-4 complexes
CC   and alter their transactivation activity. Compounds which interact with
CC   the Tcf-3 and Tcf-4 proteins and interfere with their transcriptional
CC   activation, identified using these screens, may be used to treat fetal
CC   cancer, cancers with stem cell phenotypes, testis or ovary teratomas,
CC   myeloblastomas, medulloblastomas, mammary or intestinal cancers. The
CC   present sequence represents a mouse Tcf-3 (mTcf-3) nucleotide sequence.
XX
SQ   Sequence 1945 BP; 428 A; 656 C; 507 G; 354 T; 0 other;


  Query Match             78.0%;  Score 15.6;  DB 20;  Length 1945;
  Best Local Similarity   83.3%;  Pred. No. 1.4e+02;
  Matches   15;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgasca 18
          |||:|||| ||||||:||
Db   1844 ttggtcaatatttgacca 1861


RESULT 18
F22305
ID   F22305 standard; DNA; 1082138 BP.
XX
AC   F22305;
XX
DT   20-MAR-2001  (first entry)
XX
DE   Arabidopsis thaliana chromosome 4 centromere.
XX
KW   Centromere; michrosome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200055325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US07392.
XX
PR   18-MAR-1999;   99US-0125219.
PR   01-APR-1999;   99US-0127409.
PR   18-MAY-1999;   99US-0134770.
PR   13-SEP-1999;   99US-0153584.
PR   17-SEP-1999;   99US-0154603.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D,  Copenhaver G,  Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited michrosomes which can serve as vectors for
PT   the construction of transgenic plant and animal cells
XX
PS   Claim 68; Page 977-1388; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful for
CC   producing stably inherited michrosomes which can serve as vectors for
CC   the construction of transgenic plant and animal cells expressing
CC   selected proteins such as hormones, enzymes, interleukins, clotting
CC   factors, cytokines, antibodies, and growth factors.
XX
SQ   Sequence 1082138 BP; 348775 A; 194404 C; 195515 G; 343444 T; 0 other;


  Query Match             78.0%;  Score 15.6;  DB 21;  Length 1082138;
  Best Local Similarity   93.8%;  Pred. No. 2e+02;
  Matches   15;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 tcaaaatttgascaaa 20
          ||||||||||:||||
Db 468680 tcaaaatttgagcaaa 468695


RESULT 19
Z29907/c
ID   Z29907 standard; cDNA; 829 BP.
XX
AC   Z29907;
XX
DT   26-JAN-2000  (first entry)
XX
DE   cDNA encoding a SC21 protein of soybean.
XX
KW   Promoter; seed coat specificity; SC21 protein;
KW   herbicide resistance gene; viral coat protein; biological control;
KW   Bt toxin; seed taste; ss.
XX
OS   Glycine max.
XX
PN   WO9953067-A2.
XX
PD   21-OCT-1999.
```

XX
PF 13-APR-1999; 99WO-CA00293.
XX
PR 13-APR-1998; 98US-0059090.
XX
PA (MIAC) CANADA MIN AGRIC & AGRI-FOOD CANADA.
XX
PI Miki B, Gijzen M, Miller S, Bowman L, Batchelor A, Hu M;
PI Boutilier K;
XX
DR WPI; 1999-611304/52.
XX
PT Novel promoter sequences and genes useful for inducing expression of
PT genes in plant seed coats -
XX
PS Example; Page 134; 155pp; English.
XX
CC The present sequence encodes a SC21 protein. The promoter of
CC this gene is differentially expressed in seed coat tissues,
CC specifically within the outer integument, the inner integument, the
CC thick walled parenchyma, the thin walled parenchyma, the endothelium,
CC the hourglass cells, the palisade, the stellate parenchyma, or the
CC membranous endocarp associated with the seed coat. The seed-coat
CC promoters may be used to express proteins of interest in seed coat
CC tissues. Genes of interest include but are not restricted to herbicide
CC resistance genes, genes encoding viral coat proteins, or genes encoding
CC proteins conferring biological control of pests or pathogens, e.g. a Bt
CC toxin. Other genes that may be expressed include proteins that alter the
CC taste of the seed and/or affect the nutritive value of the seed.
XX
SQ Sequence 829 BP; 200 A; 203 C; 137 G; 289 T; 0 other;

Query Match 77.0%; Score 15.4; DB 20; Length 829;
Best Local Similarity 84.2%; Pred. No. 1.6e+02;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgstcaaaatttgascaaa 20
|| |:| ||||||||| ||||
Db 438 TGGTAAAAATTTGAACAAA 420

RESULT 20
Z33573
ID Z33573 standard; cDNA; 2720 BP.
XX
AC Z33573;
XX
DT 08-DEC-1999 (first entry)
XX
DE Human breast tumour-associated EST 33.
XX
KW Expressed sequence tag; EST; human; breast; cancer; cytostatic;
KW medicaments; gene therapy; treatment; fat metabolism; ss.
XX
OS Homo sapiens.
XX
PN DE19813835-A1.
XX
PD 23-SEP-1999.
XX
PF 20-MAR-1998; 98DE-1013835.
XX
PR 20-MAR-1998; 98DE-1013835.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
XX
DR WPI; 1999-528979/45.
XX
PT Human nucleic acid sequences and protein products from normal breast
PT tissue, useful for breast cancer therapy -
XX
PS Claim 1a; 121; 206pp; German.
XX
CC This invention describes novel human nucleic acid sequences from normal
CC breast tissue which have cytostatic activity. The nucleic acid sequences
CC can be used to produce and isolate full-length gene sequences. They can
CC be used to express proteins, which can be used as tools to find an
CC activity against breast cancer. The sequences can be used in sense or
CC antisense form. They are especially useful for medicaments for gene
CC therapy to treat breast cancer and for treating illnesses associated
CC with fat metabolism. Z33541-Z33610 represent expressed sequence tags
CC described in the method of the invention.
XX
SQ Sequence 2720 BP; 947 A; 398 C; 454 G; 921 T; 0 other;

Query Match 77.0%; Score 15.4; DB 20; Length 2720;
Best Local Similarity 84.2%; Pred. No. 1.8e+02;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tgstcaaaatttgascaaa 20
|| ||||| |||||:|||
Db 371 tgatcaaaattttgagcaaa 389

Search completed: June 7, 2001, 18:09:55
Job time: 3453 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 7, 2001, 16:52:22 ; Search time 112.59 Seconds
(without alignments)
31.016 Million cell updates/sec

Title: US-09-601-561-1
Perfect score: 20
Sequence: 1 ttgstcaaaatttgascaaa 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	15	75.0	1083	4	US-09-276-531-30	Sequence 30, Appl
	2	15	75.0	2040	2	US-08-821-355A-1	Sequence 1, Appli
	3	15	75.0	2040	2	US-09-003-687A-1	Sequence 1, Appli
	4	15	75.0	2040	4	US-09-136-605-1	Sequence 1, Appli
	5	15	75.0	2444	2	US-08-821-355A-2	Sequence 2, Appli
	6	15	75.0	2444	2	US-09-003-687A-2	Sequence 2, Appli
	7	15	75.0	2444	4	US-09-136-605-2	Sequence 2, Appli
	8	14.8	74.0	2593	1	US-08-728-956-1	Sequence 1, Appli
	9	14.8	74.0	2678	1	US-08-724-194-2	Sequence 2, Appli
	10	14.8	74.0	6671	1	US-08-280-443-1	Sequence 1, Appli
	11	14.8	74.0	6671	1	US-08-457-459-1	Sequence 1, Appli
	12	14.8	74.0	6671	1	US-08-555-678-1	Sequence 1, Appli
	13	14.8	74.0	6671	5	PCT-US95-02275-1	Sequence 1, Appli
	14	14.6	73.0	1366	3	US-08-883-526-2	Sequence 2, Appli
c	15	14.6	73.0	4320	2	US-08-472-534-4	Sequence 4, Appli
c	16	14.6	73.0	6395	2	US-08-982-956-1	Sequence 1, Appli
c	17	14.6	73.0	6395	4	US-09-228-317-1	Sequence 1, Appli
	18	14.6	73.0	6578	1	US-08-514-975B-1	Sequence 1, Appli
	19	14.6	73.0	6578	5	PCT-US95-12507-1	Sequence 1, Appli
	20	14.6	73.0	15222	2	US-08-801-898A-23	Sequence 23, Appl
	21	14.6	73.0	15223	2	US-08-892-403A-1	Sequence 1, Appli
	22	14.4	72.0	888	4	US-09-188-930-13	Sequence 13, Appl
	23	14.4	72.0	2254	1	US-08-153-848-27	Sequence 27, Appl
	24	14.4	72.0	2254	4	US-09-299-843A-27	Sequence 27, Appl
	25	14.4	72.0	2254	5	PCT-US93-11153-27	Sequence 27, Appl
c	26	14.4	72.0	2256	1	US-08-181-271A-2	Sequence 2, Appli
c	27	14.4	72.0	2256	1	US-08-449-315-2	Sequence 2, Appli
c	28	14.4	72.0	2256	1	US-08-444-803-2	Sequence 2, Appli
c	29	14.4	72.0	2256	1	US-08-449-043-2	Sequence 2, Appli
c	30	14.4	72.0	2256	1	US-08-456-265A-2	Sequence 2, Appli
c	31	14.4	72.0	2256	1	US-08-455-416-2	Sequence 2, Appli
c	32	14.4	72.0	2256	1	US-08-455-244-2	Sequence 2, Appli
c	33	14.4	72.0	2256	1	US-08-454-876-2	Sequence 2, Appli
c	34	14.4	72.0	2256	2	US-08-457-364-2	Sequence 2, Appli
c	35	14.4	72.0	2256	2	US-08-456-262-2	Sequence 2, Appli
c	36	14.4	72.0	2256	2	US-08-456-240-2	Sequence 2, Appli
c	37	14.4	72.0	2256	2	US-08-455-736-2	Sequence 2, Appli
c	38	14.4	72.0	2256	2	US-08-971-217-2	Sequence 2, Appli
	39	14.4	72.0	2510	4	US-08-894-324A-2	Sequence 2, Appli
c	40	14.4	72.0	2875	3	US-08-458-434A-4	Sequence 4, Appli
	41	14.4	72.0	3119	4	US-09-299-843A-31	Sequence 31, Appl
	42	14.4	72.0	10660	2	US-08-267-803B-8	Sequence 8, Appli
c	43	14.4	72.0	15144	3	US-08-458-434A-6	Sequence 6, Appli
c	44	14.4	72.0	152331	4	US-09-128-155-16	Sequence 16, Appl
c	45	14.2	71.0	3213	2	US-08-975-527-2	Sequence 2, Appli
	46	14	70.0	161	3	US-09-004-113-6	Sequence 6, Appli
c	47	14	70.0	687	1	US-08-073-807A-11	Sequence 11, Appl
	48	14	70.0	2169	4	US-08-981-803-14	Sequence 14, Appl
c	49	14	70.0	2169	4	US-08-981-803-28	Sequence 28, Appl
	50	14	70.0	2934	1	US-08-198-446B-16	Sequence 16, Appl

ALIGNMENTS

```
RESULT   1
US-09-276-531-30
; Sequence 30, Application US/09276531
; Patent No. 6183968
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Yue, Henry
;    APPLICANT:  Reddy, Roopa
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Baughn, Mariah R.
;    TITLE OF INVENTION:  COMPOSITION FOR THE DETECTION OF GENES ENCODING
;    TITLE OF INVENTION:  RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION
;    NUMBER OF SEQUENCES:  134
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  INCYTE PHARMACEUTICALS, INC.
;      STREET:  3174 PORTER DRIVE
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Word Perfect 6.1 for Windows/MS-DOS 6.2
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/276,531
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/079,677
;      FILING DATE:  March 27, 1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lynn E. Murry, Ph.D.
;      REGISTRATION NUMBER:  42,918
;      REFERENCE/DOCKET NUMBER:  PA-0008 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 855-0555
;      TELEFAX:  (650) 845-4166
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1083 base pairs
```

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  COLNTUT02
;      CLONE:  615769
US-09-276-531-30


  Query Match             75.0%;  Score 15;  DB 4;  Length 1083;
  Best Local Similarity   78.9%;  Pred. No. 64;
  Matches   15;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 tgstcaaaatttgascaaa 20
          ||: ||||||| ||:||||
Db   1016 TGGGCAAAATTGGACCAAA 1034


RESULT   2
US-08-821-355A-1
; Sequence 1, Application US/08821355A
; Patent No. 5851775
;  GENERAL INFORMATION:
;    APPLICANT:  Barker, Nick
;    APPLICANT:  Clevers, Hans
;    APPLICANT:  Korinek, Vladimir
;    APPLICANT:  Morin, Patrice
;    APPLICANT:  Kinzler, Kenneth
;    APPLICANT:  Vogelstein, Bert
;    APPLICANT:  Sparks, Andrew
;    TITLE OF INVENTION:  Beta Catenin, TCF-4, and APC
;    TITLE OF INVENTION:  Interact to Prevent Cancer
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner & Witcoff, Ltd.
;      STREET:  1001 G Street, N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20001
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/821,355A
;      FILING DATE:  20-MAR-1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kagan, Sarah A
;      REGISTRATION NUMBER:  32,145
;      REFERENCE/DOCKET NUMBER:  1107.05064
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-508-9100
;      TELEFAX:  202-508-9299
;      TELEX:  97430 BMB UT
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2040 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-821-355A-1


  Query Match             75.0%;  Score 15;  DB 2;  Length 2040;
  Best Local Similarity   78.9%;  Pred. No. 69;
  Matches   15;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaa 19
          |||:|||| ||||||:| |
Db   1814 TTGGTCAATATTTGACCCA 1832


RESULT   3
US-09-003-687A-1
; Sequence 1, Application US/09003687A
; Patent No. 5998600
;  GENERAL INFORMATION:
;    APPLICANT:  Barker, Nick
;    APPLICANT:  Clevers, Hans
;    APPLICANT:  Korinek, Vladimir
;    APPLICANT:  Morin, Patrice
;    APPLICANT:  Kinzler, Kenneth
;    APPLICANT:  Vogelstein, Bert
;    APPLICANT:  Sparks, Andrew
;    TITLE OF INVENTION:  Beta Catenin, TCF-4, and APC
;    TITLE OF INVENTION:  Interact to Prevent Cancer
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner & Witcoff, Ltd.
;      STREET:  1001 G Street, N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20001
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/003,687A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/821,355
;      FILING DATE:  20-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kagan, Sarah A
;      REGISTRATION NUMBER:  32,145
;      REFERENCE/DOCKET NUMBER:  1107.05064
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-508-9100
;      TELEFAX:  202-508-9299
;      TELEX:  97430 BMB UT
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2040 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-003-687A-1


  Query Match             75.0%;  Score 15;  DB 2;  Length 2040;
  Best Local Similarity   78.9%;  Pred. No. 69;
  Matches   15;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaa 19
          |||:|||| ||||||:| |
Db   1814 TTGGTCAATATTTGACCCA 1832


RESULT   4
US-09-136-605-1
; Sequence 1, Application US/09136605A
; Patent No. 6140052
; GENERAL INFORMATION:
;  APPLICANT: He, Tong-Chuan
;  APPLICANT: Kinzler, Kenneth
;  APPLICANT: Vogelstein, Bert
```

```
;  TITLE OF INVENTION: Beta Catenin, TCF-4, and APC Interact to
;  TITLE OF INVENTION: Prevent Cancer
;  FILE REFERENCE: 1107.75741
;  CURRENT APPLICATION NUMBER: US/09/136,605A
;  CURRENT FILING DATE: 1998-08-20
;  EARLIER APPLICATION NUMBER: 08/821,355
;  EARLIER FILING DATE: 1997-03-20
;  EARLIER APPLICATION NUMBER: 09/003,687
;  EARLIER FILING DATE: 1998-01-06
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 2040
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-136-605-1


  Query Match             75.0%;  Score 15;  DB 4;  Length 2040;
  Best Local Similarity   78.9%;  Pred. No. 69;
  Matches   15;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ttgstcaaaatttgascaa 19
           |||:|||| ||||||:| |
Db    1814 ttggtcaatatttgaccca 1832


RESULT   5
US-08-821-355A-2
; Sequence 2, Application US/08821355A
; Patent No. 5851775
;  GENERAL INFORMATION:
;    APPLICANT:  Barker, Nick
;    APPLICANT:  Clevers, Hans
;    APPLICANT:  Korinek, Vladimir
;    APPLICANT:  Morin, Patrice
;    APPLICANT:  Kinzler, Kenneth
;    APPLICANT:  Vogelstein, Bert
;    APPLICANT:  Sparks, Andrew
;    TITLE OF INVENTION:  Beta Catenin, TCF-4, and APC
;    TITLE OF INVENTION:  Interact to Prevent Cancer
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner & Witcoff, Ltd.
;      STREET:  1001 G Street, N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20001
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/821,355A
;      FILING DATE:  20-MAR-1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kagan, Sarah A
;      REGISTRATION NUMBER:  32,145
;      REFERENCE/DOCKET NUMBER:  1107.05064
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-508-9100
;      TELEFAX:  202-508-9299
;      TELEX:  97430 BMB UT
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2444 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-821-355A-2


  Query Match             75.0%;  Score 15;  DB 2;  Length 2444;
  Best Local Similarity   78.9%;  Pred. No. 70;
  Matches   15;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ttgstcaaaatttgascaa 19
           |||:|||| ||||||:| |
Db    2245 TTGGTCAATATTTGACCCA 2263


RESULT   6
US-09-003-687A-2
; Sequence 2, Application US/09003687A
; Patent No. 5998600
;  GENERAL INFORMATION:
;    APPLICANT:  Barker, Nick
;    APPLICANT:  Clevers, Hans
;    APPLICANT:  Korinek, Vladimir
;    APPLICANT:  Morin, Patrice
;    APPLICANT:  Kinzler, Kenneth
;    APPLICANT:  Vogelstein, Bert
;    APPLICANT:  Sparks, Andrew
;    TITLE OF INVENTION:  Beta Catenin, TCF-4, and APC
;    TITLE OF INVENTION:  Interact to Prevent Cancer
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner & Witcoff, Ltd.
;      STREET:  1001 G Street, N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20001
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/003,687A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/821,355
;      FILING DATE:  20-MAR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kagan, Sarah A
;      REGISTRATION NUMBER:  32,145
;      REFERENCE/DOCKET NUMBER:  1107.05064
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-508-9100
;      TELEFAX:  202-508-9299
;      TELEX:  97430 BMB UT
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2444 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-003-687A-2


  Query Match             75.0%;  Score 15;  DB 2;  Length 2444;
  Best Local Similarity   78.9%;  Pred. No. 70;
  Matches   15;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ttgstcaaaatttgascaa 19
           |||:|||| ||||||:| |
Db    2245 TTGGTCAATATTTGACCCA 2263
```

```
RESULT   7
US-09-136-605-2
; Sequence 2, Application US/09136605A
; Patent No. 6140052
; GENERAL INFORMATION:
;  APPLICANT: He, Tong-Chuan
;  APPLICANT: Kinzler, Kenneth
;  APPLICANT: Vogelstein, Bert
;  TITLE OF INVENTION: Beta Catenin, TCF-4, and APC Interact to
;  TITLE OF INVENTION: Prevent Cancer
;  FILE REFERENCE: 1107.75741
;  CURRENT APPLICATION NUMBER: US/09/136,605A
;  CURRENT FILING DATE: 1998-08-20
;  EARLIER APPLICATION NUMBER: 08/821,355
;  EARLIER FILING DATE: 1997-03-20
;  EARLIER APPLICATION NUMBER: 09/003,687
;  EARLIER FILING DATE: 1998-01-06
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 2444
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-136-605-2


  Query Match             75.0%;  Score 15;  DB 4;  Length 2444;
  Best Local Similarity   78.9%;  Pred. No. 70;
  Matches   15;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaa 19
          |||:|||| ||||||:| |
Db   2245 ttggtcaatatttgaccca 2263


RESULT   8
US-08-728-956-1
; Sequence 1, Application US/08728956
; Patent No. 5677175
;  GENERAL INFORMATION:
;    APPLICANT:  Hodges, Thomas K.
;    APPLICANT:  McGee, J. D.
;    TITLE OF INVENTION:  Plant Pathogen Induced Proteins
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Barnes & Thornburg
;      STREET:  11 S Meridian
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46204
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/728,956
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Breen, John P.
;      REGISTRATION NUMBER:  38,833
;      REFERENCE/DOCKET NUMBER:  3220-26119
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)231-7745
;      TELEFAX:  (317)231-7433
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2593 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Oryza sativa
US-08-728-956-1


  Query Match             74.0%;  Score 14.8;  DB 1;  Length 2593;
  Best Local Similarity   80.0%;  Pred. No. 89;
  Matches   16;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||:||||||||   ||||
Db   1748 TTGGTCAAAATTTAGTCAAA 1767


RESULT   9
US-08-724-194-2
; Sequence 2, Application US/08724194
; Patent No. 5824875
;  GENERAL INFORMATION:
;    APPLICANT:  RANU, RAJINDER S.
;    TITLE OF INVENTION:  ONE-AMINOCYCLOPROPANE-1-CARBOXYLATE
;    TITLE OF INVENTION:  SYNTHASE GENES FROM PELARGONIUM TO CONTROL ETHYLENE LEVELS
;    TITLE OF INVENTION:  IN GERANIUMS
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SANTANGELO LAW OFFICES PC
;      STREET:  315 WEST OAK STREET, STE 701
;      CITY:  FORT COLLINS
;      STATE:  CO
;      COUNTRY:  USA
;      ZIP:  80521
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/724,194
;      FILING DATE:  01-OCT-1996
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SANTANGELO, LUKE
;      REGISTRATION NUMBER:  31,997
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (970) 224-3100
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2678 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-724-194-2


  Query Match             74.0%;  Score 14.8;  DB 1;  Length 2678;
  Best Local Similarity   80.0%;  Pred. No. 89;
  Matches   16;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||:|||||||  || ||||
Db   2182 TTGGTCAAAATGGGATCAAA 2201


RESULT  10
US-08-280-443-1
; Sequence 1, Application US/08280443
; Patent No. 5643778
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Nishikura, Kazuko
;    TITLE OF INVENTION:  RNA Editing Enzyme and Methods of Use
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Cntr, P.O. Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/280,443
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/197,794
;      FILING DATE:  17-FEB-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  WST49AUSA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9206
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6671 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  155..3832
US-08-280-443-1


  Query Match             74.0%;  Score 14.8;  DB 1;  Length 6671;
  Best Local Similarity   80.0%;  Pred. No. 99;
  Matches   16;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||  ||||| ||||:||||
Db   5641 TTGACCAAAAATTGACCAAA 5660


RESULT  11
US-08-457-459-1
; Sequence 1, Application US/08457459
; Patent No. 5677428
;  GENERAL INFORMATION:
;    APPLICANT:  Nishikura, Kazuko
;    TITLE OF INVENTION:  RNA Editing Enzyme and Methods of Use
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Cntr, P.O. Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,459
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/197,794
;      FILING DATE:  17-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/280,443
;      FILING DATE:  25-JUL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  WST49CUSA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9206
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6671 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  155..3832
US-08-457-459-1


  Query Match             74.0%;  Score 14.8;  DB 1;  Length 6671;
  Best Local Similarity   80.0%;  Pred. No. 99;
  Matches   16;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||  ||||| ||||:||||
Db   5641 TTGACCAAAAATTGACCAAA 5660


RESULT  12
US-08-555-678-1
; Sequence 1, Application US/08555678
; Patent No. 5763174
;  GENERAL INFORMATION:
;    APPLICANT:  Nishikura, Kazuko
;    TITLE OF INVENTION:  RNA Editing Enzyme and Methods
;    TITLE OF INVENTION:  of Use Thereof
;    NUMBER OF SEQUENCES:  67
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Cntr, P.O. Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/555,678
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/197,794
;      FILING DATE:  17-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/280,443
;      FILING DATE:  25-JUL-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,459
;      FILING DATE:  01-JUN-1995
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  WST49DUSA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9206
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6671 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  155..3832
US-08-555-678-1


  Query Match             74.0%;  Score 14.8;  DB 1;  Length 6671;
  Best Local Similarity   80.0%;  Pred. No. 99;
  Matches   16;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||  ||||| ||||:||||
Db   5641 TTGACCAAAAATTGACCAAA 5660


RESULT  13
PCT-US95-02275-1
; Sequence 1, Application PC/TUS9502275
;  GENERAL INFORMATION:
;    APPLICANT:  Wistar Institute of Anatomy &, Biology
;    TITLE OF INVENTION:  RNA Editing Enzyme and Methods of Use
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Howson and Howson
;      STREET:  Spring House Corporate Cntr, P.O. Box 457
;      CITY:  Spring House
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19477
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/02275
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/280,443
;      FILING DATE:  25-JUL-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/197,794
;      FILING DATE:  17-FEB-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bak, Mary E.
;      REGISTRATION NUMBER:  31,215
;      REFERENCE/DOCKET NUMBER:  WST49BPCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-540-9206
;      TELEFAX:  215-540-5818
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6671 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  155..3832
PCT-US95-02275-1


  Query Match             74.0%;  Score 14.8;  DB 5;  Length 6671;
  Best Local Similarity   80.0%;  Pred. No. 99;
  Matches   16;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||  ||||| ||||:||||
Db   5641 TTGACCAAAAATTGACCAAA 5660


RESULT  14
US-08-883-526-2
; Sequence 2, Application US/08883526
; Patent No. 6033893
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  NEW HUMAN CATHEPSIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/883,526
;      FILING DATE:  Herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0331 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1366 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  THYMNOT02
;      CLONE:  347021
US-08-883-526-2


  Query Match             73.0%;  Score 14.6;  DB 3;  Length 1366;
  Best Local Similarity   93.3%;  Pred. No. 1e+02;
  Matches   14;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 caaaatttgascaaa 20
          ||||||||||:||||
Db    115 CAAAATTTGACCAAA 129
```

```
RESULT 15
US-08-472-534-4/c
; Sequence 4, Application US/08472534
; Patent No. 5919620
;  GENERAL INFORMATION:
;    APPLICANT:  Hamel, Josee
;    APPLICANT:  Brodeur, Bernard R
;    APPLICANT:  Martin, Denis
;    TITLE OF INVENTION:  HEAT SHOCK PROTEIN HSP72 FROM
;    TITLE OF INVENTION:  STREPTOCOCCUS PNEUMONIAE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Neave
;      STREET:  1251 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10020
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/472,534
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haley Jr, James F
;      REGISTRATION NUMBER:  27,794
;      REFERENCE/DOCKET NUMBER:  Biovac-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-596-9000
;      TELEFAX:  212-596-9090
;      TELEX:  14-8367
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4320 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptococcus pneumoniae
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3265..4320
;      OTHER INFORMATION:  /product= "NH2-terminal portion of
;      OTHER INFORMATION:  DNA J"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  682..2502
;      OTHER INFORMATION:  /product= "Heat-Shock Protein 72"
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  682..2502
US-08-472-534-4


  Query Match             73.0%;  Score 14.6;  DB 2;  Length 4320;
  Best Local Similarity   82.4%;  Pred. No. 1.2e+02;
  Matches   14;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 stcaaaatttgascaaa 20
           :|||||||||| :||||
Db     565 CTCAAAATTTGCGCAAA 549


RESULT 16
US-08-982-956-1/c
; Sequence 1, Application US/08982956
; Patent No. 5861312
;  GENERAL INFORMATION:
;    APPLICANT:  Varshavsky, Alexander
;    APPLICANT:  Kwon, Yong Tae
;    TITLE OF INVENTION:  NUCLEIC ACID ENCODING MAMMALIAN UBR1
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kevin M. Farrell, P.C.
;      STREET:  P.O. Box 999
;      CITY:  York Harbor
;      STATE:  ME
;      COUNTRY:  US
;      ZIP:  03911
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/982,956
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farrell, Kevin M.
;      REGISTRATION NUMBER:  35,505
;      REFERENCE/DOCKET NUMBER:  CIT-2001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (207) 363-0558
;      TELEFAX:  (207) 363-0528
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6395 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  115..5385
US-08-982-956-1


  Query Match             73.0%;  Score 14.6;  DB 2;  Length 6395;
  Best Local Similarity   82.4%;  Pred. No. 1.2e+02;
  Matches   14;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 gstcaaaatttgascaa 19
           |:|||||| ||||:|||
Db    3816 GGTCAAAAGTTGAGCAA 3800


RESULT 17
US-09-228-317-1/c
; Sequence 1, Application US/09228317
; Patent No. 6159732
;  GENERAL INFORMATION:
;    APPLICANT:  Varshavsky, Alexander
;    APPLICANT:  Kwon, Yong Tae
;    TITLE OF INVENTION:  NUCLEIC ACID ENCODING MAMMALIAN UBR1
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kevin M. Farrell, P.C.
;      STREET:  P.O. Box 999
;      CITY:  York Harbor
;      STATE:  ME
;      COUNTRY:  US
;      ZIP:  03911
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/228,317
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farrell, Kevin M.
;      REGISTRATION NUMBER:  35,505
;      REFERENCE/DOCKET NUMBER:  CIT-2001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (207) 363-0558
;      TELEFAX:  (207) 363-0528
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6395 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  115..5385
US-09-228-317-1


  Query Match             73.0%;  Score 14.6;  DB 4;  Length 6395;
  Best Local Similarity   82.4%;  Pred. No. 1.2e+02;
  Matches   14;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 gstcaaaatttgascaa 19
          |:|||||| ||||:|||
Db   3816 GGTCAAAAGTTGAGCAA 3800


RESULT  18
US-08-514-975B-1
; Sequence 1, Application US/08514975B
; Patent No. 5789229
;  GENERAL INFORMATION:
;    APPLICANT:  Dr. Gail W. Wertz et al.
;    TITLE OF INVENTION:  Gene Therapy Vectors and Vaccines
;    TITLE OF INVENTION:  Based on No. 5789229-Segmented Negative Stranded RNA Viruses
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dr. Benjamin A. Adler
;      STREET:  8011 Candle Lane
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple
;      OPERATING SYSTEM:  Macintosh 7.5.3
;      SOFTWARE:  Macintosh Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/514,975B
;      FILING DATE:  September 29, 1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/316,438
;      FILING DATE:  30-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adler Ph.D., Benjamin A.
;      REGISTRATION NUMBER:  35,423
;      REFERENCE/DOCKET NUMBER:  D5899CIP2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713-777-2321
;      TELEFAX:  713-777-6908
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6578 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  9..6504
US-08-514-975B-1


  Query Match             73.0%;  Score 14.6;  DB 1;  Length 6578;
  Best Local Similarity   82.4%;  Pred. No. 1.2e+02;
  Matches   14;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 tgstcaaaatttgasca 18
          ||:| |||||||||:||
Db   5714 TGGTGAAAATTTGACCA 5730


RESULT  19
PCT-US95-12507-1
; Sequence 1, Application PC/TUS9512507
;  GENERAL INFORMATION:
;    APPLICANT:  Dr. Gail W. Wertz et al.
;    TITLE OF INVENTION:  Prevention and Treatment of Respiratory Tract      Disease
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109-1875
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/12507
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/316,438
;      FILING DATE:  30-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Arnold, Beth E. (BEA)
;      REGISTRATION NUMBER:  35,430
;      REFERENCE/DOCKET NUMBER:  UAG-010CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)227-5941
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6578 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  9..6504
PCT-US95-12507-1


  Query Match             73.0%;  Score 14.6;  DB 5;  Length 6578;
  Best Local Similarity   82.4%;  Pred. No. 1.2e+02;
  Matches   14;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 tgstcaaaatttgasca 18
          ||:| |||||||||:||
Db   5714 TGGTGAAAATTTGACCA 5730


RESULT  20
```

```
US-08-801-898A-23
; Sequence 23, Application US/08801898A
; Patent No. 5998602
; GENERAL INFORMATION:
;   APPLICANT: Torrence, Paul F.
;   APPLICANT: Silverman, Robert H.
;   APPLICANT: Cirino, Nick M.
;   APPLICANT: Li, Guiying
;   APPLICANT: Xiao, Wei
;   TITLE OF INVENTION: RNASE L ACTIVATORS AND ANTISENSE
;   TITLE OF INVENTION: OLIGONUCLEOTIDES EFFECTIVE TO TREAT RSV INFECTIONS
;   NUMBER OF SEQUENCES: 29
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Pennie & Edmonds LLP
;     STREET: 1155 Avenue of the Americas
;     CITY: New York
;     STATE: NY
;     COUNTRY: U.S.A.
;     ZIP: 10036
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette
;     COMPUTER: IBM Compatible
;     OPERATING SYSTEM: DOS
;     SOFTWARE: FastSEQ Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/801,898A
;     FILING DATE: 18-FEB-1997
;     CLASSIFICATION: 514
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Poissant, Brian M.
;     REGISTRATION NUMBER: 28,462
;     REFERENCE/DOCKET NUMBER: 8656-009
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 212-790-9090
;     TELEFAX: 212-869-9741
;     TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 23:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 15222 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA
;   FEATURE:
;     NAME/KEY: RSV-A2
;     LOCATION: 1...15222
;     OTHER INFORMATION:
US-08-801-898A-23


  Query Match            73.0%;  Score 14.6;  DB 2;  Length 15222;
  Best Local Similarity  82.4%;  Pred. No. 1.4e+02;
  Matches   14;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 tgstcaaaatttgasca 18
          ||:| |||||||||:||
Db  14203 TGGTGAAAATTTGACCA 14219


Search completed: June 7, 2001, 18:06:02
Job time: 4420 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 7, 2001, 16:12:02 ; Search time 1600.48 Seconds
(without alignments)
109.168 Million cell updates/sec

Title: US-09-601-561-1
Perfect score: 20
Sequence: 1 ttgstcaaaatttgascaaa 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_gss_fun:*
177: em_gss_hum1:*
178: em_gss_hum2:*
179: em_gss_hum3:*
180: em_gss_hum4:*
181: em_gss_hum5:*
182: em_gss_hum6:*
183: em_gss_hum7:*
184: em_gss_hum8:*
185: em_gss_hum9:*
186: em_gss_inv1:*
187: em_gss_inv2:*
188: em_gss_inv3:*
189: em_gss_other:*
190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	17.6	88.0	422	142	BE984433	BE984433 UI-M-CG0p
c	2	17.6	88.0	443	139	BE805957	BE805957 ss51h03.y
	3	17.6	88.0	531	142	BE983494	BE983494 UI-M-CG0p
	4	17.6	88.0	544	142	BE981494	BE981494 UI-M-CG0p
	5	17.6	88.0	628	122	AW944186	AW944186 SD01538.3
c	6	17.6	88.0	687	202	AQ108788	AQ108788 CIT-HSP-2
	7	17.6	88.0	834	142	BE972776	BE972776 601652006
c	8	17.2	86.0	401	214	AQ954031	AQ954031 nbeb0053J
	9	17.2	86.0	854	151	BF669926	BF669926 602118430
c	10	17	85.0	283	127	BB198261	BB198261 BB198261
	11	16.6	83.0	182	115	AW443329	AW443329 EST308259
c	12	16.6	83.0	317	167	BE445953	BE445953 WHE1143_A
	13	16.6	83.0	355	113	AW266160	AW266160 L30-2853T
	14	16.6	83.0	370	136	BE577343	BE577343 L48-2171T
c	15	16.6	83.0	375	12	AA856224	AA856224 L30-247T3
	16	16.6	83.0	461	8	AA537391	AA537391 vj98a09.r
c	17	16.6	83.0	475	112	AW220819	AW220819 EST297288
c	18	16.6	83.0	480	167	BE444676	BE444676 WHE1137_F

c	19	16.6	83.0	533	110	AW034294	AW034294 EST277865
	20	16.6	83.0	543	221	AZ401935	AZ401935 1M0169F03
	21	16.6	83.0	545	165	BE330861	BE330861 so86h01.y
	22	16.6	83.0	547	120	AW830687	AW830687 sm06e04.y
c	23	16.6	83.0	548	225	AZ634875	AZ634875 1M0490J23
c	24	16.6	83.0	574	110	AW033917	AW033917 EST277488
	25	16.6	83.0	575	146	BF266468	BF266468 HV_CEa001
c	26	16.6	83.0	579	24	AI770998	AI770998 SWAMCAC31
	27	16.6	83.0	622	111	AW134165	AW134165 fi15d07.y
	28	16.6	83.0	630	225	AZ636272	AZ636272 1M0495H06
c	29	16.6	83.0	665	136	BE576819	BE576819 L0-1625T3
	30	16.6	83.0	697	162	BE036290	BE036290 MO15A04 M
	31	16.6	83.0	955	231	CNS04V6W	AL308705 Tetraodon
	32	16.6	83.0	957	232	CNS051WY	AL317419 Tetraodon
c	33	16.4	82.0	437	215	AZ046379	AZ046379 nbeb0085N
	34	16.4	82.0	446	228	B98792	B98792 CIT-HSP-228
	35	16.4	82.0	450	204	AQ263042	AQ263042 CITBI-E1-
	36	16.4	82.0	460	107	AU166766	AU166766 AU166766
c	37	16.4	82.0	480	219	AZ301611	AZ301611 Bng60/F c
c	38	16.4	82.0	552	207	AQ503227	AQ503227 V34E07 mT
c	39	16.4	82.0	552	212	AQ872695	AQ872695 V34E7 mTn
c	40	16.4	82.0	558	103	AI896675	AI896675 EST266118
	41	16.4	82.0	601	31	AV676198	AV676198 AV676198
	42	16.4	82.0	674	218	AZ270212	AZ270212 RPCI-23-4
	43	16.4	82.0	795	206	AQ396074	AQ396074 nbxb0066F
c	44	16.4	82.0	807	212	AQ864440	AQ864440 nbeb0023J
	45	16.4	82.0	885	141	BE896894	BE896894 601439408
	46	16.2	81.0	484	209	AQ664451	AQ664451 HS_5480_B
	47	16.2	81.0	781	29	AV383190	AV383190 AV383190
c	48	16.2	81.0	852	232	CNS05ENF	AL333924 Tetraodon
c	49	16	80.0	243	11	AA750707	AA750707 ISGS0295
	50	16	80.0	252	122	BB011557	BB011557 BB011557

ALIGNMENTS

RESULT 1
BE984433
LOCUS BE984433 422 bp mRNA EST 05-OCT-2000
DEFINITION UI-M-CG0p-bfy-d-12-0-UI.s1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone UI-M-CG0p-bfy-d-12-0-UI 3', mRNA sequence.
ACCESSION BE984433
VERSION BE984433.1 GI:10656659
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 422)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
Thé sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the retina tissue cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements The following repetitive elements were found in this cDNA sequence: 171-318, >Lx3#LINE/L1
316-416, >Lx6#LINE/L1
Seq primer: M13 Forward
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .422
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-CG0p-bfy-d-12-0-UI"
/clone_lib="NIH_BMAP_Ret4_S2"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_Ret4_S2 library is a subtracted library, ultimately derived from mouse retina tissue libraries at various stages of development. For a detailed description of the library from which this clone was derived, please visit our web site at brainest.eng.uiowa.edu.
TAG_LIB=NIH_BMAP_Ret4_S2
TAG_TISSUE=adult-retina
TAG_SEQ=GTCAGCGCGCAC"
BASE COUNT 103 a 88 c 84 g 146 t 1 others
ORIGIN

Query Match 88.0%; Score 17.6; DB 142; Length 422;
Best Local Similarity 85.0%; Pred. No. 2.4e+02;
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 ttgstcaaaatttgascaaa 20
          |||:||||||||||:|| |
Db    116 TTGCTCAAAATTTGACCACA 135
```

RESULT 2
BE805957/c
LOCUS BE805957 443 bp mRNA EST 20-SEP-2000
DEFINITION ss51h03.y1 Gm-c1061 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1061-2286 5' similar to TR:O24657 O24657 DAL1 PROTEIN. [1] ;, mRNA sequence.
ACCESSION BE805957
VERSION BE805957.1 GI:10237069
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 443)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann ,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
High quality sequence stop: 426.
FEATURES Location/Qualifiers
source 1. .443

```
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1061-2286"
                     /clone_lib="Gm-c1061"
                     /tissue_type="mature flowers of field grown plants"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     from mature flowers of field grown plants for the cultivar
                     Raiden. Complementary DNA was synthesized from mRNA using
                     a primer consisting of a poly(dT) sequence with a XhoI
                     restriction site. EcoRI adapters were ligated to the
                     blunt-ended cDNA fragments followed by XhoI digestion. The
                     cDNA fragments  were directionally cloned into the
                     EcoRI-XhoI restriction site of the pBluescript vector. The
                     ligated cDNA fragments were transformed into DH10B host
                     cells (GibcoBRL). This library was constructed in the
                     laboratory of Dr. Randy Shoemaker."
BASE COUNT      124 a     76 c    112 g    131 t
ORIGIN


  Query Match             88.0%;  Score 17.6;  DB 139;  Length 443;
  Best Local Similarity   85.0%;  Pred. No. 2.5e+02;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||:||||||||||:|| |
Db    265 TTGGTCAAAATTTGAGCATA 246


RESULT   3
BE983494
LOCUS       BE983494      531 bp    mRNA            EST       05-OCT-2000
DEFINITION  UI-M-CG0p-bfx-d-03-0-UI.s1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone
            UI-M-CG0p-bfx-d-03-0-UI 3', mRNA sequence.
ACCESSION   BE983494
VERSION     BE983494.1  GI:10654734
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            retina tissue cDNA Library Preparation: M.B. Soares Lab Clone
            distribution: Researchers may obtain BMAP cDNA clones from RESEARCH
            GENETICS. It should be noted that Bento Soares is generating a
            small number of additional specialized non-redundant arrays of BMAP
            cDNAs whose availability will be considered under appropriate and
            limited collaborative arrangements The following repetitive
            elements were found in this cDNA sequence: 171-318, >Lx3#LINE/L1
            316-512, >Lx6#LINE/L1
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .531
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-CG0p-bfx-d-03-0-UI"
                     /clone_lib="NIH_BMAP_Ret4_S2"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_Ret4_S2 library is a subtracted library,
                     ultimately derived from mouse retina tissue libraries at
                     various stages of development. For a detailed description
                     of the library from which this clone was derived,  please
                     visit our web site at brainest.eng.uiowa.edu.
                     TAG_LIB=NIH_BMAP_Ret4_S2
                     TAG_TISSUE=adult-retina
                     TAG_SEQ=GTCAGCGCGCAC"
BASE COUNT      124 a    109 c    120 g    178 t
ORIGIN


  Query Match             88.0%;  Score 17.6;  DB 142;  Length 531;
  Best Local Similarity   85.0%;  Pred. No. 2.5e+02;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||:||||||||||:|| |
Db    116 TTGCTCAAAATTTGACCACA 135


RESULT   4
BE981494
LOCUS       BE981494      544 bp    mRNA            EST       05-OCT-2000
DEFINITION  UI-M-CG0p-bdc-g-11-0-UI.s1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone
            UI-M-CG0p-bdc-g-11-0-UI 3', mRNA sequence.
ACCESSION   BE981494
VERSION     BE981494.1  GI:10650656
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            retina tissue cDNA Library Preparation: M.B. Soares Lab Clone
            distribution: Researchers may obtain BMAP cDNA clones from RESEARCH
            GENETICS. It should be noted that Bento Soares is generating a
            small number of additional specialized non-redundant arrays of BMAP
            cDNAs whose availability will be considered under appropriate and
            limited collaborative arrangements The following repetitive
            elements were found in this cDNA sequence: 171-318, >Lx3#LINE/L1
            316-544, >Lx6#LINE/L1
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .544
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
```

```
                     /db_xref="taxon:10090"
                     /clone="UI-M-CG0p-bdc-g-11-0-UI"
                     /clone_lib="NIH_BMAP_Ret4_S2"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_Ret4_S2 library is a subtracted library,
                     ultimately derived from mouse retina tissue libraries at
                     various stages of development. For a detailed description
                     of the library from which this clone was derived, please
                     visit our web site at brainest.eng.uiowa.edu.
                     TAG_LIB=NIH_BMAP_Ret4_S2
                     TAG_TISSUE=adult-retina
                     TAG_SEQ=GTCAGCGCGCAC"
BASE COUNT      131 a    103 c    132 g    178 t
ORIGIN


  Query Match             88.0%;  Score 17.6;  DB 142;  Length 544;
  Best Local Similarity   85.0%;  Pred. No. 2.5e+02;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||:|||||||||||:|| |
Db    116 TTGCTCAAAATTTGACCACA 135


RESULT   5
AW944186
LOCUS       AW944186      628 bp    mRNA            EST       31-MAY-2000
DEFINITION  SD01538.3prime SD Drosophila melanogaster Schneider L2 cell culture
            pOT2 Drosophila melanogaster cDNA clone SD01538 3, mRNA sequence.
ACCESSION   AW944186
VERSION     AW944186.1  GI:8121914
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 628)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
            ,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Other_ESTs: SD01538.5prime
            Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Based upon the presence of a XhoI site followed by a run of 14 or
            more T residues at the beginning of the sequence, this clone was
            polyadenylated. The resulting Poly-T sequence has been removed.
            Plate: SD.15  row: D  column: 2
            High quality sequence stop: 492.
FEATURES             Location/Qualifiers
     source          1. .628
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="SD01538"
                     /clone_lib="SD Drosophila melanogaster Schneider L2 cell
                     culture pOT2"
                     /lab_host="DH5-alpha"
                     /note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
                     fractionated cDNAs were directly ligated into pOT2.
                     Plasmid cDNA library."
BASE COUNT      212 a     91 c     78 g    246 t      1 others
ORIGIN


  Query Match             88.0%;  Score 17.6;  DB 122;  Length 628;
  Best Local Similarity   85.0%;  Pred. No. 2.6e+02;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||:|||||||||||: |||
Db    572 TTGCTCAAAATTTGAGAAAA 591


RESULT   6
AQ108788/c
LOCUS       AQ108788      687 bp    DNA             GSS       29-AUG-1998
DEFINITION  CIT-HSP-2372G13.TR CIT-HSP Homo sapiens genomic clone 2372G13, DNA
            sequence.
ACCESSION   AQ108788
VERSION     AQ108788.1  GI:3485478
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 687)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2372G13.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .687
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2372G13"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT      234 a    121 c    125 g    207 t
ORIGIN


  Query Match             88.0%;  Score 17.6;  DB 202;  Length 687;
  Best Local Similarity   85.0%;  Pred. No. 2.6e+02;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||:||||||| ||||:||||
Db    673 TTGCTCAAAAGTTGAGCAAA 654


RESULT   7
BE972776
LOCUS       BE972776      834 bp    mRNA            EST       04-OCT-2000
DEFINITION  601652006F1 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:3935377 5',
            mRNA sequence.
ACCESSION   BE972776
VERSION     BE972776.1  GI:10586112
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 834)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM777  row: 1  column: 02
            High quality sequence stop: 543.
FEATURES             Location/Qualifiers
     source          1. .834
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3935377"
                     /clone_lib="NIH_MGC_82"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).   Average insert size
                     1.35 kb (range 0.9-4.0 kb).  14/15 colonies contained
                     inserts by PCR. This library was enriched for full-length
                     clones and was constructed by Clontech  Laboratories (Palo
                     Alto, CA)."
BASE COUNT      237 a    151 c    191 g    255 t
ORIGIN


  Query Match             88.0%;  Score 17.6;  DB 142;  Length 834;
  Best Local Similarity   85.0%;  Pred. No. 2.7e+02;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaaa 20
          |||:|| ||||||||:||||
Db    731 TTGGTCTAAATTTGAGCAAA 750


RESULT   8
AQ954031/c
LOCUS       AQ954031      401 bp    DNA             GSS       27-JAN-2000
DEFINITION  nbeb0053J14r CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
            clone OSJNBb0053J14r, DNA sequence.
ACCESSION   AQ954031
VERSION     AQ954031.1  GI:6777296
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
            Oryza.
REFERENCE   1  (bases 1 to 401)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: GCTATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence stop: 328.
FEATURES             Location/Qualifiers
     source          1. .401
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="OSJNBb0053J14r"
                     /clone_lib="CUGI Rice BAC Library (EcoRI)"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
                     Rice is the most important food crop in the world. Half of
                     the world population, especially those inhabiting highly
                     populated areas of the humid tropics and subtropics, rely
                     on rice as their primary source of carbohydrate.
                     Monocotyledonous rice is a diploid plant (2n=24) with a
                     haploid genome equivalent of 431 Mbp (Arumuganathan and
                     Earle, 1991). The relatively small genome of rice, three
                     times larger than that of Arabidopsis, makes it suitable
                     for genomic studies. In order to facilitate positional
                     cloning, physical mapping and genome sequencing of rice,
                     we have constructed a BAC library from Oryza sativa,
                     Nipponbare variety using EcoRI as the cloning enzyme. The
                     library contains 55,296 clones with an average insert size
                     of 121 Kb providing approximatley 15 haploid genome
                     equivalents. The deep coverage allows the isolation a
                     particular sequence with a probability of 99.9 %. Three
                     high density filters, each containing 18,432 clones
                     (doubly spotted), represent the whole library for colony
                     screening and can be requested from the Clemson University
                     BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT      142 a     66 c     62 g    130 t      1 others
ORIGIN


  Query Match             86.0%;  Score 17.2;  DB 214;  Length 401;
  Best Local Similarity   88.9%;  Pred. No. 3.8e+02;
  Matches   16;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 tgstcaaaatttgascaa 19
          ||:||||||||||:|||
Db     44 TGCTCAAAATTTGAGCAA 27


RESULT   9
BF669926
LOCUS       BF669926      854 bp    mRNA            EST       21-DEC-2000
DEFINITION  602118430F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4275606 5',
            mRNA sequence.
ACCESSION   BF669926
VERSION     BF669926.1  GI:11943821
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 854)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1094  row: 1  column: 07
```

```
          High quality sequence stop: 575.
FEATURES             Location/Qualifiers
     source          1. .854
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4275606"
                     /clone_lib="NIH_MGC_56"
                     /tissue_type="primitive neuroectoderm"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA.  5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.65 kb (range 0.9-4.0 kb).  15/15 colonies
                     contained  inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      264 a    139 c    172 g    279 t
ORIGIN


  Query Match             86.0%;  Score 17.2;  DB 151;  Length 854;
  Best Local Similarity   88.9%;  Pred. No. 4.2e+02;
  Matches   16;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gstcaaaatttgascaaa 20
           |:||||||||||:||||
Db     445 GGTCAAAATTTGAGCAAA 462


RESULT  10
BB198261/c
LOCUS       BB198261      283 bp    mRNA            EST       30-JUN-2000
DEFINITION  BB198261 RIKEN full-length enriched, adult male spinal cord Mus
            musculus cDNA clone A330108P03 3', mRNA sequence.
ACCESSION   BB198261
VERSION     BB198261.1  GI:8863214
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 283)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .283
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="A330108P03"
                     /clone_lib="RIKEN full-length enriched, adult male spinal
                     cord"
                     /sex="male"
                     /tissue_type="spinal cord"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. "
BASE COUNT       97 a     54 c     65 g     67 t
ORIGIN


  Query Match             85.0%;  Score 17;  DB 127;  Length 283;
  Best Local Similarity   89.5%;  Pred. No. 4.5e+02;
  Matches   17;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 tgstcaaaatttgascaaa 20
           || ||||||||||:||||
Db     248 TGTTCAAAATTTGACCAAA 230


RESULT  11
AW443329
LOCUS       AW443329      182 bp    mRNA            EST       17-FEB-2000
DEFINITION  EST308259 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA
            clone cLET43H8 5', mRNA sequence.
ACCESSION   AW443329
VERSION     AW443329.1  GI:6985511
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
            I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 182)
  AUTHORS   D'Ascenzo,M., He,X., Lyman,J., Matern,A.L., Vision,T., Holt,I.E.,
            Liang,F., Hansen,T.S., Ronning,C.M., Craven,M.B., Bowman,C.L.,
            Nierman,W., Fraser,C.M., Venter,J.C., Tanksley,S.D., Giovannoni
            ,J.J. and Martin,G.B.
  TITLE     Generation of ESTs from tomato callus (mixed elicitor)
  JOURNAL   Unpublished (1999)
```

```
COMMENT     Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU
            5 prime sequence.
FEATURES             Location/Qualifiers
     source          1. .182
                     /organism="Lycopersicon esculentum"
                     /cultivar="Rio Grande PtoR"
                     /db_xref="taxon:4081"
                     /clone="cLET43H8"
                     /clone_lib="tomato mixed elicitor, BTI"
                     /tissue_type="leaf"
                     /dev_stage="4-6 week old plants"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; cLET - Inoculated with a variety of disease response
                     elicitors. Plants  exposed to 2,6 dichloroisonicotinic
                     acid, BTH, jasmonic acid, ethylene,  fenthion, EIX,
                     okadaic acid, or systemin prior to tissue harvest.  EcoRI
                     site was destroyed during cloning."
BASE COUNT       66 a     39 c     38 g     39 t
ORIGIN


  Query Match             83.0%;  Score 16.6;  DB 115;  Length 182;
  Best Local Similarity   84.2%;  Pred. No. 6.6e+02;
  Matches   16;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 tgstcaaaatttgascaaa 20
          ||:||||||| |||:||||
Db    163 TGGTCAAAATGTGAGCAAA 181


RESULT  12
BE445953/c
LOCUS       BE445953      317 bp    mRNA            EST       25-JUL-2000
DEFINITION  WHE1143_A07_B13ZS Wheat etiolated seedling root normalized cDNA
            library Triticum aestivum cDNA clone WHE1143_A07_B13, mRNA
            sequence.
ACCESSION   BE445953
VERSION     BE445953.1  GI:9445515
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae;
            Triticum.
REFERENCE   1  (bases 1 to 317)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
            ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Nguyen,H.T.,
            Rausch,C.J., Seaton,C.L., Tong,J.C. and Zhang,D.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Normalized root cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .317
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE1143_A07_B13"
                     /clone_lib="Wheat etiolated seedling root normalized cDNA
                     library"
                     /tissue_type="Root"
                     /dev_stage="Five day old etiolated seedling"
                     /lab_host="E. coli DH10B"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid
                     pBluescript SK; Site_1: EcoRI; Site_2: XhoI; Seeds were
                     surface-sterilized, germinated and grown aseptically in
                     the  dark at room temperature on filter paper with water,
                     nystatin and  cefotaxime in covered crystallization
                     dishes. Roots were harvested.  The tissue, total RNA, and
                     poly(A) RNA were prepared, a cDNA library  was made in the
                     TJ Close lab (Choi, Close, Fenton) at the University  of
                     California, Riverside.  The cDNA clones were in vivo
                     excised to  give pBluescript phagemids before
                     normalization was carried out.  The mass excision of
                     phagemid library and normalization were done in HT  Nguyen
                     lab by D. Zhang at Texas Tech Univeristy.  Normalization
                     protocol used was that of Soares.  Plasmid DNA
                     preparations and DNA  sequencing were performed in the OD
                     Anderson lab (all other authors)."
BASE COUNT       73 a     75 c     62 g     107 t
ORIGIN


  Query Match             83.0%;  Score 16.6;  DB 167;  Length 317;
  Best Local Similarity   84.2%;  Pred. No. 7.1e+02;
  Matches   16;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ttgstcaaaatttgascaa 19
            |||:||| |||||||:|||
Db    255 TTGGTCATAATTTGAGCAA 237


RESULT  13
AW266160
LOCUS       AW266160      355 bp    mRNA            EST       29-DEC-1999
DEFINITION  L30-2853T3 Ice plant Lambda Uni-Zap XR expression library, 30 hours
            NaCl treatment Mesembryanthemum crystallinum cDNA clone L30-2853
            5', mRNA sequence.
ACCESSION   AW266160
VERSION     AW266160.1  GI:6644847
KEYWORDS    EST.
SOURCE      common ice plant.
  ORGANISM  Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;
            Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1  (bases 1 to 355)
  AUTHORS   Cushman, J.C.
  TITLE     An expressed sequence tag database for the common ice plant,
            Mesembryanthemum crystallinum
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Cushman JC
            Department of Biochemistry and Molecular Biology
            Oklahoma State University
            350 Noble Research Center, Stillwater, OK 74078-3035, USA
            Tel: 405-744-6207
            Fax: 405-744-7799
            Email: jcushman@biochem.okstate.edu
            Poly(A) tail, 22 nt: 356. .377
            PCR PRimers
            FORWARD: T7
            BACKWARD: T3
            Insert Length: 377   Std Error: 0.00
            Plate: L30-29  row: E  column: 9
            Seq primer: T3
            High quality sequence stop: 355.
FEATURES             Location/Qualifiers
     source          1. .355
                     /organism="Mesembryanthemum crystallinum"
                     /db_xref="taxon:3544"
```

/clone="L30-2853"
/clone_lib="Ice plant Lambda Uni-Zap XR expression library
, 30 hours NaCl treatment"
/tissue_type="Leaf, 30 h 0.4M NaCl"
/dev_stage="Six week old"
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
BASE COUNT 102 a 47 c 84 g 122 t
ORIGIN

Query Match 83.0%; Score 16.6; DB 113; Length 355;
Best Local Similarity 84.2%; Pred. No. 7.2e+02;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ttgstcaaaatttgascaa 19
|||:| ||||||||||:|||
Db 3 TTGGTTAAAATTTGAGCAA 21

RESULT 14
BE577343
LOCUS BE577343 370 bp mRNA EST 15-AUG-2000
DEFINITION L48-2171T3 Ice plant Lambda Uni-Zap XR expression library, 48 hours NaCl treatment Mesembryanthemum crystallinum cDNA clone L48-2171 5', mRNA sequence.
ACCESSION BE577343
VERSION BE577343.1 GI:9827142
KEYWORDS EST.
SOURCE common ice plant.
ORGANISM Mesembryanthemum crystallinum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE 1 (bases 1 to 370)
AUTHORS Cushman, J.C.
TITLE An expressed sequence tag database for the common ice plant, Mesembryanthemum crystallinum
JOURNAL Unpublished (1997)
COMMENT Contact: Cushman JC
Department of Biochemistry and Molecular Biology
Oklahoma State University
350 Noble Research Center, Stillwater, OK 74078-3035, USA
Tel: 405-744-6207
Fax: 405-744-7799
Email: jcushman@biochem.okstate.edu
PCR PRimers
FORWARD: T7
BACKWARD: T3
Plate: L48-22 row: F column: 11
Seq primer: T3
High quality sequence stop: 350
POLYA=No.
FEATURES Location/Qualifiers
source 1. .370
/organism="Mesembryanthemum crystallinum"
/db_xref="taxon:3544"
/clone="L48-2171"
/clone_lib="Ice plant Lambda Uni-Zap XR expression library
, 48 hours NaCl treatment"
/tissue_type="Leaf, 48 h 0.4M NaCl"
/dev_stage="Six week old"
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
BASE COUNT 107 a 49 c 86 g 128 t
ORIGIN

Query Match 83.0%; Score 16.6; DB 136; Length 370;
Best Local Similarity 84.2%; Pred. No. 7.3e+02;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ttgstcaaaatttgascaa 19
|||:| ||||||||||:|||
Db 20 TTGGTTAAAATTTGAGCAA 38

RESULT 15
AA856224/c
LOCUS AA856224 375 bp mRNA EST 09-MAR-1998
DEFINITION L30-247T3 Ice plant Lambda Uni-Zap XR expression library, 30 hours NaCl treatment Mesembryanthemum crystallinum cDNA clone L30-247 5', mRNA sequence.
ACCESSION AA856224
VERSION AA856224.1 GI:2944526
KEYWORDS EST.
SOURCE common ice plant.
ORGANISM Mesembryanthemum crystallinum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE 1 (bases 1 to 375)
AUTHORS Cushman, J.C.
TITLE An expressed sequence tag database for the common ice plant, Mesembryanthemum crystallinum
JOURNAL Unpublished (1997)
COMMENT Contact: Cushman JC
Department of Biochemistry and Molecular Biology
Oklahoma State University
350 Noble Research Center, Stillwater, OK 74078-3035, USA
Tel: 405-744-6207
Fax: 405-744-7799
Email: jcushman@biochem.okstate.edu
PCR PRimers
FORWARD: T7
BACKWARD: T3
Insert Length: 447 Std Error: 10.00
Plate: L30-3 row: E column: 3
Seq primer: T3
High quality sequence stop: 300.
FEATURES Location/Qualifiers
source 1. .375
/organism="Mesembryanthemum crystallinum"
/db_xref="taxon:3544"
/clone="L30-247"
/clone_lib="Ice plant Lambda Uni-Zap XR expression library
, 30 hours NaCl treatment"
/tissue_type="Leaf, 30 h 0.4M NaCl"
/dev_stage="Six week old"
/note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
BASE COUNT 128 a 85 c 49 g 113 t
ORIGIN

Query Match 83.0%; Score 16.6; DB 12; Length 375;
Best Local Similarity 84.2%; Pred. No. 7.3e+02;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ttgstcaaaatttgascaa 19
|||:| ||||||||||:|||
Db 359 TTGGTTAAAATTTGAGCAA 341

RESULT 16
AA537391
LOCUS AA537391 461 bp mRNA EST 29-JUL-1997
DEFINITION vj98a09.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA clone IMAGE:945112 5', mRNA sequence.
ACCESSION AA537391
VERSION AA537391.1 GI:2283384
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:541968
            High quality sequence stop: 313.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Mus musculus"
                     /strain="B6D2 F1/J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:945112"
                     /clone_lib="Knowles Solter mouse blastocyst B1"
                     /tissue_type="blastocyst"
                     /dev_stage="embryo (pre-implantation)"
                     /lab_host="DH10B"
                     /note="Organ: embryo; Vector: pSPORT; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally from mRNA prepared
                     from 800 blastocysts.  Primer: SalI(dT):
                     5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTTT-3'.  cDNAs  were
                     cloned into the NotI/SalI sites of a pSPORT vector (Life
                     Technologies).  Two different size selections:  B1 (larger
                     inserts) and B3."
BASE COUNT       98 a    121 c    142 g    100 t
ORIGIN


  Query Match             83.0%;  Score 16.6;  DB 8;  Length 461;
  Best Local Similarity   84.2%;  Pred. No. 7.5e+02;
  Matches   16;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 tgstcaaaatttgascaaa 20
           ||:|||||||||||:| ||
Db     411 TGCTCAAAATTTGACCTAA 429


RESULT  17
AW220819/c
LOCUS       AW220819      475 bp    mRNA            EST       07-DEC-1999
DEFINITION  EST297288 tomato fruit mature green, TAMU Lycopersicon esculentum
            cDNA clone cLEF2D7, mRNA sequence.
ACCESSION   AW220819
VERSION     AW220819.1  GI:6532503
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
            I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 475)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
            Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M.
            , Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
  TITLE     Generation of ESTs from tomato fruit tissue
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU
            5 prime sequence.
FEATURES             Location/Qualifiers
     source          1. .475
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEF2D7"
                     /clone_lib="tomato fruit mature green, TAMU"
                     /tissue_type="fruit pericarp"
                     /dev_stage="mature green (3-5 days pre-ripening)"
                     /lab_host="SOLR"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; cLEF - Fruit were tagged at the 1cm stage and
                     harvested 3-5 days prior to  ripening. Fruit were cut in
                     half to verify the seeds were indeed  'immature' and the
                     seeds and locules were discarded prior to freezing the
                     pericarp"
BASE COUNT      110 a    103 c     87 g    175 t
ORIGIN


  Query Match             83.0%;  Score 16.6;  DB 112;  Length 475;
  Best Local Similarity   84.2%;  Pred. No. 7.5e+02;
  Matches   16;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 ttgstcaaaatttgascaa 19
           |||:||| |||||||:|||
Db     288 TTGGTCATAATTTGAGCAA 270


RESULT  18
BE444676/c
LOCUS       BE444676      480 bp    mRNA            EST       25-JUL-2000
DEFINITION  WHE1137_F11_K21ZS Wheat etiolated seedling root normalized cDNA
            library Triticum aestivum cDNA clone WHE1137_F11_K21, mRNA
            sequence.
ACCESSION   BE444676
VERSION     BE444676.1  GI:9444226
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae;
            Triticum.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
            ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Nguyen,H.T.,
            Rausch,C.J., Seaton,C.L., Tong,J.C. and Zhang,D.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Normalized root cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE1137_F11_K21"
                     /clone_lib="Wheat etiolated seedling root normalized cDNA
                     library"
                     /tissue_type="Root"
                     /dev_stage="Five day old etiolated seedling"
```

/lab_host="E. coli DH10B"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid pBluescript SK; Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized, germinated and grown aseptically in the dark at room temperature on filter paper with water, nystatin and cefotaxime in covered crystallization dishes. Roots were harvested. The tissue, total RNA, and poly(A) RNA were prepared, a cDNA library was made in the TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. The cDNA clones were in vivo excised to give pBluescript phagemids before normalization was carried out. The mass excision of phagemid library and normalization were done in HT Nguyen lab by D. Zhang at Texas Tech Univeristy. Normalization protocol used was that of Soares. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT 106 a 114 c 105 g 155 t
ORIGIN

Query Match 83.0%; Score 16.6; DB 167; Length 480;
Best Local Similarity 84.2%; Pred. No. 7.5e+02;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ttgstcaaaatttgascaa 19
||||:||| ||||||||:|||
Db 377 TTGGTCATAATTTGAGCAA 359

RESULT 19
AW034294/c
LOCUS AW034294 533 bp mRNA EST 15-SEP-1999
DEFINITION EST277865 tomato callus, TAMU Lycopersicon esculentum cDNA clone cLEC37H6, mRNA sequence.
ACCESSION AW034294
VERSION AW034294.1 GI:5893050
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 533)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE Generation of ESTs from tomato callus tissue
JOURNAL Unpublished (1999)
COMMENT Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
FEATURES Location/Qualifiers
source 1. .533
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEC37H6"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XL1-Blue MRF'"
/note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2: Xho1; supplier: Giovannoni laboratory; cLEC - Cotyledons of seedlings 7-10 days post-germination were excised, cut at both ends and placed on MS medium with no selection. Mixed callus was harvested at 25 and 40 days and included undifferentiated masses. Tomato Callus EST Library"

BASE COUNT 126 a 115 c 100 g 192 t
ORIGIN

Query Match 83.0%; Score 16.6; DB 110; Length 533;
Best Local Similarity 84.2%; Pred. No. 7.6e+02;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ttgstcaaaatttgascaa 19
||||:||| ||||||||:|||
Db 355 TTGGTCATAATTTGAGCAA 337

RESULT 20
AZ401935
LOCUS AZ401935 543 bp DNA GSS 03-OCT-2000
DEFINITION 1M0169F03F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0169F03 F, DNA sequence.
ACCESSION AZ401935
VERSION AZ401935.1 GI:10517009
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 543)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0169 row: F column: 03
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 543.
FEATURES Location/Qualifiers
source 1. .543
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0169F03"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."
BASE COUNT 167 a 112 c 110 g 154 t
ORIGIN

Query Match 83.0%; Score 16.6; DB 221; Length 543;
Best Local Similarity 84.2%; Pred. No. 7.7e+02;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ttgstcaaaatttgascaa 19
||||:||||||| |||:|||
Db 328 TTGCTCAAAATATGAGCAA 346

Search completed: June 7, 2001, 17:41:28
Job time: 5366 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 7, 2001, 18:04:28 ; Search time 1335.93 Seconds
(without alignments)
463.655 Million cell updates/sec

Title: US-09-601-561-2
Perfect score: 42
Sequence: 1 tattctttgctcaaaatttg.....caaattttgagcaaagaata 42

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		25.2	60.0	622	53	CNS01NIY	AL152283 Anopheles
c 2		25.2	60.0	622	53	CNS01NIY	AL152283 Anopheles
3		25	59.5	163733	77	AC083911	AC083911 Rattus no
c 4		25	59.5	163733	77	AC083911	AC083911 Rattus no
5		24.6	58.6	268238	86	AC008372	AC008372 Homo sapi
c 6		24.6	58.6	268238	86	AC008372	AC008372 Homo sapi
7		24.4	58.1	161733	70	AC025797	AC025797 Homo sapi
c 8		24.4	58.1	161733	70	AC025797	AC025797 Homo sapi
9		24.4	58.1	177099	81	AL442067	AL442067 Homo sapi
c 10		24.4	58.1	177099	81	AL442067	AL442067 Homo sapi
11		24.4	58.1	179640	86	AC006322	AC006322 Homo sapi

c	12	24.4	58.1	179640	86	AC006322
	13	24.4	58.1	183285	75	AC073498
c	14	24.4	58.1	183285	75	AC073498
	15	24.4	58.1	212307	75	AC073609
c	16	24.4	58.1	212307	75	AC073609
	17	24.2	57.6	104913	92	HSJ718P11
c	18	24.2	57.6	104913	92	HSJ718P11
	19	24.2	57.6	126413	76	AC078977
c	20	24.2	57.6	126413	76	AC078977
	21	24.2	57.6	192748	62	AC011089
c	22	24.2	57.6	192748	62	AC011089
	23	24.2	57.6	193004	70	AC025907
c	24	24.2	57.6	193004	70	AC025907
	25	24	57.1	1523	88	AF156775
c	26	24	57.1	1523	88	AF156775
	27	24	57.1	1660	88	AF156774
c	28	24	57.1	1660	88	AF156774
	29	24	57.1	14922	6	CELY55F3AR
c	30	24	57.1	14922	6	CELY55F3AR
	31	24	57.1	42254	5	CELC44B12
c	32	24	57.1	42254	5	CELC44B12
	33	24	57.1	65147	77	AC084003
c	34	24	57.1	65147	77	AC084003
	35	24	57.1	78153	6	CEY53C10A
c	36	24	57.1	78153	6	CEY53C10A
	37	24	57.1	110000	83	CEY48G10_1
c	38	24	57.1	110000	83	CEY48G10_1
	39	24	57.1	110000	83	CEY53C10_2
c	40	24	57.1	110000	83	CEY53C10_2
	41	24	57.1	123091	6	CEY87G2A
c	42	24	57.1	123091	6	CEY87G2A
	43	24	57.1	126615	85	AC005180
c	44	24	57.1	126615	85	AC005180
	45	24	57.1	145549	68	AC023115
c	46	24	57.1	145549	68	AC023115
	47	24	57.1	145744	61	AC010414
c	48	24	57.1	145744	61	AC010414
	49	24	57.1	147726	82	AP000684
c	50	24	57.1	147726	82	AP000684

AC006322 Homo sapi
AC073498 Homo sapi
AC073498 Homo sapi
AC073609 Mus muscu
AC073609 Mus muscu
AL109983 Human DNA
AL109983 Human DNA
AC078977 Oryza sat
AC078977 Oryza sat
AC011089 Homo sapi
AC011089 Homo sapi
AC025907 Oryza sat
AC025907 Oryza sat
AF156775 Homo sapi
AF156775 Homo sapi
AF156774 Homo sapi
AF156774 Homo sapi
AC024827 Caenorhab
AC024827 Caenorhab
AF036692 Caenorhab
AF036692 Caenorhab
AC084003 Homo sapi
AC084003 Homo sapi
AL033536 Caenorhab
AL033536 Caenorhab
Continuation (2 of
Continuation (2 of
Continuation (3 of
Continuation (3 of
AL110500 Caenorhab
AL110500 Caenorhab
AC005180 Homo sapi
AC005180 Homo sapi
AC023115 Homo sapi
AC023115 Homo sapi
AC010414 Homo sapi
AC010414 Homo sapi
AP000684 Homo sapi
AP000684 Homo sapi

ALIGNMENTS

```
RESULT   1
CNS01NIY
LOCUS       CNS01NIY      622 bp    DNA             STS       18-FEB-2000
DEFINITION  Anopheles gambiae STS T7 end of clone 24C24 of NotreDame1 library
            from strain PEST of Anopheles gambiae (African malaria mosquito),
            sequence tagged site.
ACCESSION   AL152283
VERSION     AL152283.1  GI:7013202
KEYWORDS    STS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
            Culicoidea; Culicidae; Anopheles.
REFERENCE   1  (bases 1 to 622)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
REFERENCE   2  (bases 1 to 622)
  AUTHORS   Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
            Roux, Paris 75015, France
COMMENT     This clone is from an A. gambiae BAC library provided by F.H.
            Collins and sequenced by Genoscope in collaboration with the
            Laboratory of Biochem. and Biol. Molec. of Insects, Institut
            Pasteur.
FEATURES             Location/Qualifiers
     source          1. .622
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="24C24"
                     /clone_lib="NotreDame1"
                     /note="end : T7"
BASE COUNT      221 a     97 c     95 g    187 t     22 others
ORIGIN


  Query Match           60.0%;  Score 25.2;  DB 53;  Length 622;
  Best Local Similarity   67.5%;  Pred. No. 2.8e+02;
  Matches   27;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 tattctttgctcaaaatttgatcaaattttgagcaaagaa 40
           ::|| ||:|: || || ||  |||||||||:|| ||| ||
Db     500 YWTTTTTKGYGCAGAAATTATTCAAATTTTKAGTAAAAAA 539


RESULT   2
CNS01NIY/c
LOCUS       CNS01NIY      622 bp    DNA             STS       18-FEB-2000
DEFINITION  Anopheles gambiae STS T7 end of clone 24C24 of NotreDame1 library
            from strain PEST of Anopheles gambiae (African malaria mosquito),
            sequence tagged site.
ACCESSION   AL152283
VERSION     AL152283.1  GI:7013202
KEYWORDS    STS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
            Culicoidea; Culicidae; Anopheles.
REFERENCE   1  (bases 1 to 622)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
REFERENCE   2  (bases 1 to 622)
  AUTHORS   Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
            Roux, Paris 75015, France
COMMENT     This clone is from an A. gambiae BAC library provided by F.H.
            Collins and sequenced by Genoscope in collaboration with the
            Laboratory of Biochem. and Biol. Molec. of Insects, Institut
            Pasteur.
FEATURES             Location/Qualifiers
     source          1. .622
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="24C24"
                     /clone_lib="NotreDame1"
                     /note="end : T7"
BASE COUNT      221 a     97 c     95 g    187 t     22 others
ORIGIN


  Query Match           60.0%;  Score 25.2;  DB 53;  Length 622;
  Best Local Similarity   67.5%;  Pred. No. 2.8e+02;
  Matches   27;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       3 ttctttgctcaaaatttgatcaaattttgagcaaagaata 42
           || ||| ||:|||||||||  || || || :|:|| ||::
Db     539 TTTTTTACTMAAAATTTGAATAATTTCTGCRCMAAAAAWR 500


RESULT   3
AC083911
```

```
LOCUS       AC083911   163733 bp    DNA             HTG       06-NOV-2000
DEFINITION  Rattus norvegicus clone RP32-316K5, WORKING DRAFT SEQUENCE, 23
            unordered pieces.
ACCESSION   AC083911
VERSION     AC083911.9  GI:11079382
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 163733)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
            and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163733)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 3, 2000 this sequence version replaced gi:10944432.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: TUAG
                Center clone name: RP32-316K5
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 128299 bases at least Q40
                Consensus quality: 144464 bases at least Q30
                Consensus quality: 153662 bases at least Q20
                Estimated insert size: 155201; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 3.1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 23 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    20001: contig of 20001 bp in length
            *    20002    20101: gap of unknown length
            *    20102    38094: contig of 17993 bp in length
            *    38095    38194: gap of unknown length
            *    38195    51792: contig of 13598 bp in length
            *    51793    51892: gap of unknown length
            *    51893    68031: contig of 16139 bp in length
            *    68032    68131: gap of unknown length
            *    68132    81496: contig of 13365 bp in length
            *    81497    81596: gap of unknown length
            *    81597    92431: contig of 10835 bp in length
            *    92432    92531: gap of unknown length
            *    92532   105952: contig of 13421 bp in length
            *   105953   106052: gap of unknown length
            *   106053   113442: contig of 7390 bp in length
            *   113443   113542: gap of unknown length
            *   113543   120445: contig of 6903 bp in length
            *   120446   120545: gap of unknown length
            *   120546   128253: contig of 7708 bp in length
            *   128254   128353: gap of unknown length
            *   128354   134387: contig of 6034 bp in length
            *   134388   134487: gap of unknown length
            *   134488   138270: contig of 3783 bp in length
            *   138271   138370: gap of unknown length
            *   138371   142708: contig of 4338 bp in length
            *   142709   142808: gap of unknown length
            *   142809   146428: contig of 3620 bp in length
            *   146429   146528: gap of unknown length
            *   146529   149119: contig of 2591 bp in length
            *   149120   149219: gap of unknown length
            *   149220   152397: contig of 3178 bp in length
            *   152398   152497: gap of unknown length
            *   152498   154219: contig of 1722 bp in length
            *   154220   154319: gap of unknown length
            *   154320   156024: contig of 1705 bp in length
            *   156025   156124: gap of unknown length
            *   156125   158275: contig of 2151 bp in length
            *   158276   158375: gap of unknown length
            *   158376   159884: contig of 1509 bp in length
            *   159885   159984: gap of unknown length
            *   159985   161466: contig of 1482 bp in length
            *   161467   161566: gap of unknown length
            *   161567   162596: contig of 1030 bp in length
            *   162597   162696: gap of unknown length
            *   162697   163733: contig of 1037 bp in length.
FEATURES             Location/Qualifiers
     source          1. .163733
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="RP32-316K5"
BASE COUNT    50766 a  33088 c  31597 g  46053 t   2229 others
ORIGIN
```

Query Match 59.5%; Score 25; DB 77; Length 163733;
Best Local Similarity 75.6%; Pred. No. 2.6e+02;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy      2 attctttgctcaaaatttgatcaaattttgagcaaagaata 42
          ||| | |   ||| |||| ||||||| ||  ||||||||||
Db 156915 ATTATGTTTGCAATATTTAATCAAATCTTAGGCAAAGAATA 156955

RESULT  4
AC083911/c
LOCUS       AC083911   163733 bp    DNA             HTG       06-NOV-2000
DEFINITION  Rattus norvegicus clone RP32-316K5, WORKING DRAFT SEQUENCE, 23
            unordered pieces.
ACCESSION   AC083911
VERSION     AC083911.9  GI:11079382
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 163733)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
            and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163733)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 3, 2000 this sequence version replaced gi:10944432.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: TUAG
                Center clone name: RP32-316K5
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 128299 bases at least Q40
                Consensus quality: 144464 bases at least Q30
                Consensus quality: 153662 bases at least Q20
                Estimated insert size: 155201; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 3.1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 23 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    20001: contig of 20001 bp in length
            *    20002    20101: gap of unknown length
            *    20102    38094: contig of 17993 bp in length
            *    38095    38194: gap of unknown length
            *    38195    51792: contig of 13598 bp in length
            *    51793    51892: gap of unknown length
            *    51893    68031: contig of 16139 bp in length
            *    68032    68131: gap of unknown length
            *    68132    81496: contig of 13365 bp in length
            *    81497    81596: gap of unknown length
            *    81597    92431: contig of 10835 bp in length
            *    92432    92531: gap of unknown length
            *    92532   105952: contig of 13421 bp in length
            *   105953   106052: gap of unknown length
            *   106053   113442: contig of 7390 bp in length
            *   113443   113542: gap of unknown length
            *   113543   120445: contig of 6903 bp in length
            *   120446   120545: gap of unknown length
            *   120546   128253: contig of 7708 bp in length
            *   128254   128353: gap of unknown length
            *   128354   134387: contig of 6034 bp in length
            *   134388   134487: gap of unknown length
            *   134488   138270: contig of 3783 bp in length
            *   138271   138370: gap of unknown length
            *   138371   142708: contig of 4338 bp in length
            *   142709   142808: gap of unknown length
            *   142809   146428: contig of 3620 bp in length
            *   146429   146528: gap of unknown length
            *   146529   149119: contig of 2591 bp in length
            *   149120   149219: gap of unknown length
            *   149220   152397: contig of 3178 bp in length
            *   152398   152497: gap of unknown length
            *   152498   154219: contig of 1722 bp in length
            *   154220   154319: gap of unknown length
            *   154320   156024: contig of 1705 bp in length
            *   156025   156124: gap of unknown length
            *   156125   158275: contig of 2151 bp in length
            *   158276   158375: gap of unknown length
            *   158376   159884: contig of 1509 bp in length
            *   159885   159984: gap of unknown length
            *   159985   161466: contig of 1482 bp in length
            *   161467   161566: gap of unknown length
            *   161567   162596: contig of 1030 bp in length
            *   162597   162696: gap of unknown length
            *   162697   163733: contig of 1037 bp in length.
FEATURES             Location/Qualifiers
     source          1. .163733
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
```

```
                 /clone="RP32-316K5"
BASE COUNT    50766 a  33088 c  31597 g  46053 t   2229 others
ORIGIN


  Query Match             59.5%;  Score 25;  DB 77;  Length 163733;
  Best Local Similarity   75.6%;  Pred. No. 2.6e+02;
  Matches   31;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaagaat 41
          ||||||||||  || |||||||| |||| |||   | | |||
Db 156955 TATTCTTTGCCTAAGATTTGATTAAATATTGCAAACATAAT 156915


RESULT   5
AC008372
LOCUS       AC008372   268238 bp    DNA              PRI       08-JAN-2000
DEFINITION  Homo sapiens chromosome 14 clones RP11-111O16 and RP11-61F4
            containing genes for nuclear receptor coactivator NCoA-62 (nuclear
            receptor coactivator NCoA-62) gene, complete cds; and unknown gene,
            complete sequence.
ACCESSION   AC008372
VERSION     AC008372.6  GI:6682588
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 268238)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
            Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
            James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
            and Hood,L.
  TITLE     Sequencing of human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 268238)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Dors,M., Dickhoff,R.,
            Harrison,G., James,R., Loretz,C., Lasky,S., Madan,A., Prescott,S.,
            Ratcliffe,A., Shaffer,T. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 268238)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
            Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
            James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
            and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-2000) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
COMMENT     On Jan 8, 2000 this sequence version replaced gi:6563507.
            ------------ Genome Center
                Center: Multimegabase Sequencing Center
                Center code: UWMSC
                Web site: http://chroma.mbt.washington.edu/msg_www
                Contact: leerowen@u.washington.edu
            ------------ Summary Statistics
                Sequencing vector: pUC18; L08752
                Chemistry: Dye-terminator Big Dye; 90% of reads
                Chemistry: Dye-primer Big Dye; 10% of reads
                Assembly program: Phrap; version 0.990399
                ------------
FEATURES             Location/Qualifiers
     source          1. .268238
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14q24.3"
                     /clone="RP11-111o16 and RP11-61F4"
                     /clone_lib="RPCI human BAC library 11"
                     /note="RP11-111O16 overlaps CTD-2175M1, AC008044 and
                     RP11-61F4, this entry.  RP11-61F4 overlaps RP11-63D17,
                     AL049832."
     misc_feature    1. .24249
                     /note="overlap with CTD-2175M1, AC008044"
     misc_feature    1. .186358
                     /note="Region covered by RP11-111O16"
     gene            complement(8710. .52264)
                     /gene="nuclear receptor coactivator NCoA-62"
                     /note="Intron-exon boundaries defined in relation to cDNA
                     in AF045184.  Also see U51432"
     CDS             complement(join(9199. .9397,9478. .9641,11822. .11939,
                     14292. .14388,22099. .22240,23596. .23712,26058. .26123,
                     27048. .27117,28082. .28186,29889. .29995,30077. .30172,
                     42430. .42591,46078. .46231,52225. .52238))
                     /gene="nuclear receptor coactivator NCoA-62"
                     /codon_start=1
                     /product="nuclear receptor coactivator NCoA-62"
                     /protein_id="AAF23325.1"
                     /db_xref="GI:6682589"
                     /translation="MALTSFLPAPTQLSQDQLEAEEKARSQRSRQTSLVSSRREPPPY
                     GYRKGWIPRLLEDFGDGGAFPEIHVAQYPLDMGRKKKMSNALAIQVDSEGKIKYDAIA
                     RQGQSKDKVIYSKYTDLVPKEVMNADDPDLQRPDEEAIKEITEKTRVALEKSVSQKVA
                     AAMPVRAADKLAPAQYIRYTPSQQGVAFNSGAKQRVIRMVEMQKDPMEPPRFKINKKI
                     PRGPPSPPAPVMHSPSRKMTVKEQQEWKIPPCISNWKNAKGYTIPLDKRLAADGRGLQ
                     TVHINENFAKLAEALYIADRKAREAVEMRAQVERKMAQKEKEKHEEKLREMAQKARER
                     RAGIKTHVEKEDGEARERDEIRHDRRKERQHDRNLSRAAPDKRSKLQRNENRDISEVI
                     ALGVPNPRTSNEVQYDQRLFNQSKGMDSGFAGGEDEIYNVYDQAWRGGKDMAQSIYRP
                     SKNLDKDMYGDDLEARIKTNRFVPDKEFSGSDRRQRGREGPVQFEEDPFGLDKFLEEA
                     KQHGGSKRPSDSSRPKEHEHEGKKRRKE"
     unsure          35890
                     /note="low quality data"
     unsure          45090. .45120
                     /note="low quality data"
     unsure          62460. .62510
     unsure          71720. .71770
                     /note="low quality data"
     unsure          74330. .74370
                     /note="loe quality data"
     unsure          77730. .77740
                     /note="low quality data"
     misc_feature    83558. .268238
                     /note="Regioin covered by RP11-61F4.  RP11-111O16 and
                     RP11-61F4 are the same haplotype"
     CDS             join(110090. .110224,113546. .113629,178202. .178360,
                     190211. .190369,198914. .199030,215568. .215717,
                     216875. .217072,222629. .222822,224331. .224502)
                     /note="ABC transporter-like; This new gene is defined in
                     relation to a contig of ESTs including AI118010, AW060860,
                     H30689, AI907812, AA284562 and W74413"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAF23326.1"
                     /db_xref="GI:6682590"
                     /translation="MARKALKLASWTSMALAASGIYFYSNKYLDPNDFGAVRVGRAVA
                     TTAVISYDYLTSLKSVPYGSEEYLQLRSKIHDLFQSFDDTPLGTASLAQVHKAVLHDG
                     RTVAVKVQHPKVRAQSSKDILLMEVLVLAVKQLFPEFEFMWLVDEAKKNLPLELDFLN
                     EGRNAEKVSQMLRHFDFLKVPRIHWDLSTERVLLMEFVDGGQVNDRDYMERNKIDVNE
                     ISRHLGKMYSEMIFVNGFVHCDPHPGNVLVRKHPGTGKAEIVLLDHGLYQMLTEEFRL
                     NYCHLWQSLIWTDMKRVKEYSQRLGAGDLYPLFACMLTARSWDSVNRGISQAPVTATE
                     DLEIRNNAANYLPQISHLLNHVPRQMLLILKTNDLLRGIEAALGTRASASSFLNMSRC
                     CIRALAEHKKKNTCSFFRRTQISFSEAFNLWQINLHELILRVKGLKLADRVLALICWL
                     FPAPL"
     unsure          130710. .130770
                     /note="low quality data"
     unsure          189370. .189390
                     /note="low quality data"
     unsure          191480. .191510
                     /note="low quality data"
     unsure          231590. .231630
                     /note="low quality data"
     misc_feature    250941. .268238
                     /note="overlap with RP11-63D17, AL049832"
BASE COUNT    69534 a  59016 c  63055 g  76633 t
ORIGIN
```

```
  Query Match             58.6%;  Score 24.6;  DB 86;  Length 268238;
  Best Local Similarity   76.9%;  Pred. No. 3.4e+02;
  Matches   30;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaaga 39
          | || |||||||||  ||| | ||||| || || |||||
Db 155392 TCTTGTTTGCTCAAGGTTTAAGCAAATGTTCAGAAAAGA 155430


RESULT  6
AC008372/c
LOCUS       AC008372   268238 bp    DNA             PRI       08-JAN-2000
DEFINITION  Homo sapiens chromosome 14 clones RP11-111O16 and RP11-61F4
            containing genes for nuclear receptor coactivator NCoA-62 (nuclear
            receptor coactivator NCoA-62) gene, complete cds; and unknown gene,
            complete sequence.
ACCESSION   AC008372
VERSION     AC008372.6  GI:6682588
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 268238)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
            Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
            James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
            and Hood,L.
  TITLE     Sequencing of human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 268238)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Dors,M., Dickhoff,R.,
            Harrison,G., James,R., Loretz,C., Lasky,S., Madan,A., Prescott,S.,
            Ratcliffe,A., Shaffer,T. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 268238)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
            Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
            James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
            and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-2000) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
COMMENT     On Jan 8, 2000 this sequence version replaced gi:6563507.
            ------------ Genome Center
                Center: Multimegabase Sequencing Center
                Center code: UWMSC
                Web site: http://chroma.mbt.washington.edu/msg_www
                Contact: leerowen@u.washington.edu
            ------------ Summary Statistics
                Sequencing vector: pUC18; L08752
                Chemistry: Dye-terminator Big Dye; 90% of reads
                Chemistry: Dye-primer Big Dye; 10% of reads
                Assembly program: Phrap; version 0.990399
                ------------.
FEATURES             Location/Qualifiers
     source          1. .268238
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /map="14q24.3"
                     /clone="RP11-111o16 and RP11-61F4"
                     /clone_lib="RPCI human BAC library 11"
                     /note="RP11-111O16 overlaps CTD-2175M1, AC008044 and
                     RP11-61F4, this entry.  RP11-61F4 overlaps RP11-63D17,
                     AL049832."
     misc_feature    1. .24249
                     /note="overlap with CTD-2175M1, AC008044"
     misc_feature    1. .186358
                     /note="Region covered by RP11-111O16"
     gene            complement(8710. .52264)
                     /gene="nuclear receptor coactivator NCoA-62"
                     /note="Intron-exon boundaries defined in relation to cDNA
                     in AF045184.  Also see U51432"
     CDS             complement(join(9199. .9397,9478. .9641,11822. .11939,
                     14292. .14388,22099. .22240,23596. .23712,26058. .26123,
                     27048. .27117,28082. .28186,29889. .29995,30077. .30172,
                     42430. .42591,46078. .46231,52225. .52238))
                     /gene="nuclear receptor coactivator NCoA-62"
                     /codon_start=1
                     /product="nuclear receptor coactivator NCoA-62"
                     /protein_id="AAF23325.1"
                     /db_xref="GI:6682589"
                     /translation="MALTSFLPAPTQLSQDQLEAEEKARSQRSRQTSLVSSRREPPPY
                     GYRKGWIPRLLEDFGDGGAFPEIHVAQYPLDMGRKKKMSNALAIQVDSEGKIKYDAIA
                     RQGQSKDKVIYSKYTDLVPKEVMNADDPDLQRPDEEAIKEITEKTRVALEKSVSQKVA
                     AAMPVRAADKLAPAQYIRYTPSQQGVAFNSGAKQRVIRMVEMQKDPMEPPRFKINKKI
                     PRGPPSPPAPVMHSPSRKMTVKEQQEWKIPPCISNWKNAKGYTIPLDKRLAADGRGLQ
                     TVHINENFAKLAEALYIADRKAREAVEMRAQVERKMAQKEKEKHEEKLREMAQKARER
                     RAGIKTHVEKEDGEARERDEIRHDRRKERQHDRNLSRAAPDKRSKLQRNENRDISEVI
                     ALGVPNPRTSNEVQYDQRLFNQSKGMDSGFAGGEDEIYNVYDQAWRGGKDMAQSIYRP
                     SKNLDKDMYGDDLEARIKTNRFVPDKEFSGSDRRQRGREGPVQFEEDPFGLDKFLEEA
                     KQHGGSKRPSDSSRPKEHEHEGKKRRKE"
     unsure          35890
                     /note="low quality data"
     unsure          45090. .45120
                     /note="low quality data"
     unsure          62460. .62510
     unsure          71720. .71770
                     /note="low quality data"
     unsure          74330. .74370
                     /note="loe quality data"
     unsure          77730. .77740
                     /note="low quality data"
     misc_feature    83558. .268238
                     /note="Regioin covered by RP11-61F4.  RP11-111O16 and
                     RP11-61F4 are the same haplotype"
     CDS             join(110090. .110224,113546. .113629,178202. .178360,
                     190211. .190369,198914. .199030,215568. .215717,
                     216875. .217072,222629. .222822,224331. .224502)
                     /note="ABC transporter-like; This new gene is defined in
                     relation to a contig of ESTs including AI118010, AW060860,
                     H30689, AI907812, AA284562 and W74413"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAF23326.1"
                     /db_xref="GI:6682590"
                     /translation="MARKALKLASWTSMALAASGIYFYSNKYLDPNDFGAVRVGRAVA
                     TTAVISYDYLTSLKSVPYGSEEYLQLRSKIHDLFQSFDDTPLGTASLAQVHKAVLHDG
                     RTVAVKVQHPKVRAQSSKDILLMEVLVLAVKQLFPEFEFMWLVDEAKKNLPLELDFLN
                     EGRNAEKVSQMLRHFDFLKVPRIHWDLSTERVLLMEFVDGGQVNDRDYMERNKIDVNE
                     ISRHLGKMYSEMIFVNGFVHCDPHPGNVLVRKHPGTGKAEIVLLDHGLYQMLTEEFRL
                     NYCHLWQSLIWTDMKRVKEYSQRLGAGDLYPLFACMLTARSWDSVNRGISQAPVTATE
                     DLEIRNNAANYLPQISHLLNHVPRQMLLILKTNDLLRGIEAALGTRASASSFLNMSRC
                     CIRALAEHKKKNTCSFFRRTQISFSEAFNLWQINLHELILRVKGLKLADRVLALICWL
                     FPAPL"
     unsure          130710. .130770
                     /note="low quality data"
     unsure          189370. .189390
                     /note="low quality data"
     unsure          191480. .191510
                     /note="low quality data"
     unsure          231590. .231630
                     /note="low quality data"
     misc_feature    250941. .268238
                     /note="overlap with RP11-63D17, AL049832"
BASE COUNT    69534 a  59016 c  63055 g  76633 t
ORIGIN


  Query Match             58.6%;  Score 24.6;  DB 86;  Length 268238;
  Best Local Similarity   76.9%;  Pred. No. 3.4e+02;
  Matches   30;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;
```

```
Qy      4 tctttgctcaaaatttgatcaaattttgagcaaagaata 42
          ||||| || || ||||| | |||  |||||||||| || |
Db 155430 TCTTTTCTGAACATTTGCTTAAACCTTGAGCAAACAAGA 155392


RESULT  7
AC025797
LOCUS       AC025797   161733 bp    DNA             HTG       25-JUN-2000
DEFINITION  Homo sapiens clone RP11-385C22, WORKING DRAFT SEQUENCE, 32
            unordered pieces.
ACCESSION   AC025797
VERSION     AC025797.3  GI:8705135
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161733)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-385C22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161733)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 25, 2000 this sequence version replaced gi:7534070.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L8166
                Center clone name: 385_C_22
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 139384 bases at least Q40
                Consensus quality: 151678 bases at least Q30
                Consensus quality: 156111 bases at least Q20
                Insert size: 171000; agarose-fp
                Insert size: 158633; sum-of-contigs
                Quality coverage: 3.3 in Q20 bases; agarose-fp
                Quality coverage: 3.6 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 32 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1167: contig of 1167 bp in length
            *     1168 1267: gap of      100 bp
            *     1268     2620: contig of 1353 bp in length
            *     2621 2720: gap of      100 bp
            *     2721     4325: contig of 1605 bp in length
            *     4326 4425: gap of      100 bp
            *     4426     5495: contig of 1070 bp in length
            *     5496 5595: gap of      100 bp
            *     5596     6690: contig of 1095 bp in length
            *     6691 6790: gap of      100 bp
            *     6791     8804: contig of 2014 bp in length
            *     8805 8904: gap of      100 bp
            *     8905     9390: contig of 486 bp in length
            *     9391 9490: gap of      100 bp
            *     9491    10893: contig of 1403 bp in length
            *    10894 10993: gap of      100 bp
            *    10994    12705: contig of 1712 bp in length
            *    12706 12805: gap of      100 bp
            *    12806    14698: contig of 1893 bp in length
            *    14699 14798: gap of      100 bp
            *    14799    16989: contig of 2191 bp in length
            *    16990 17089: gap of      100 bp
            *    17090    19915: contig of 2826 bp in length
            *    19916 20015: gap of      100 bp
            *    20016    23312: contig of 3297 bp in length
            *    23313 23412: gap of      100 bp
            *    23413    25456: contig of 2044 bp in length
            *    25457 25556: gap of      100 bp
            *    25557    28737: contig of 3181 bp in length
            *    28738 28837: gap of      100 bp
            *    28838    31790: contig of 2953 bp in length
            *    31791 31890: gap of      100 bp
            *    31891    36413: contig of 4523 bp in length
            *    36414 36513: gap of      100 bp
            *    36514    41435: contig of 4922 bp in length
            *    41436 41535: gap of      100 bp
            *    41536    46569: contig of 5034 bp in length
            *    46570 46669: gap of      100 bp
            *    46670    52443: contig of 5774 bp in length
            *    52444 52543: gap of      100 bp
            *    52544    58987: contig of 6444 bp in length
            *    58988 59087: gap of      100 bp
            *    59088    65801: contig of 6714 bp in length
            *    65802 65901: gap of      100 bp
            *    65902    74132: contig of 8231 bp in length
            *    74133 74232: gap of      100 bp
            *    74233    80963: contig of 6731 bp in length
            *    80964 81063: gap of      100 bp
            *    81064    87836: contig of 6773 bp in length
            *    87837 87936: gap of      100 bp
            *    87937    95449: contig of 7513 bp in length
            *    95450 95549: gap of      100 bp
            *    95550   103931: contig of 8382 bp in length
            *   103932 104031: gap of      100 bp
            *   104032   114614: contig of 10583 bp in length
            *   114615 114714: gap of      100 bp
            *   114715   126025: contig of 11311 bp in length
            *   126026 126125: gap of      100 bp
            *   126126   136164: contig of 10039 bp in length
            *   136165 136264: gap of      100 bp
            *   136265   150282: contig of 14018 bp in length
            *   150283 150382: gap of      100 bp
            *   150383   161733: contig of 11351 bp in length.
FEATURES             Location/Qualifiers
     source          1. .161733
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
                     /clone="RP11-385C22"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1167
                     /note="assembly_fragment"
     misc_feature    1268. .2620
                     /note="assembly_fragment"
     misc_feature    2721. .4325
                     /note="assembly_fragment"
     misc_feature    4426. .5495
                     /note="assembly_fragment"
     misc_feature    5596. .6690
                     /note="assembly_fragment"
     misc_feature    6791. .8804
                     /note="assembly_fragment"
     misc_feature    8905. .9390
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    9491. .10893
                     /note="assembly_fragment"
     misc_feature    10994. .12705
                     /note="assembly_fragment"
     misc_feature    12806. .14698
                     /note="assembly_fragment"
     misc_feature    14799. .16989
                     /note="assembly_fragment"
     misc_feature    17090. .19915
                     /note="assembly_fragment"
     misc_feature    20016. .23312
                     /note="assembly_fragment"
     misc_feature    23413. .25456
                     /note="assembly_fragment"
     misc_feature    25557. .28737
                     /note="assembly_fragment"
     misc_feature    28838. .31790
                     /note="assembly_fragment"
     misc_feature    31891. .36413
                     /note="assembly_fragment"
     misc_feature    36514. .41435
                     /note="assembly_fragment"
     misc_feature    41536. .46569
                     /note="assembly_fragment"
     misc_feature    46670. .52443
                     /note="assembly_fragment"
     misc_feature    52544. .58987
                     /note="assembly_fragment"
     misc_feature    59088. .65801
                     /note="assembly_fragment"
     misc_feature    65902. .74132
                     /note="assembly_fragment"
     misc_feature    74233. .80963
                     /note="assembly_fragment"
     misc_feature    81064. .87836
                     /note="assembly_fragment"
     misc_feature    87937. .95449
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    95550. .103931
                     /note="assembly_fragment"
     misc_feature    104032. .114614
                     /note="assembly_fragment"

  Query Match             58.1%;  Score 24.4;  DB 70;  Length 161733;
  Best Local Similarity   73.8%;  Pred. No. 4e+02;
  Matches   31;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 tattctttgctcaaaatttgatcaaattttgagcaaagaata 42
           || ||||| || |   ||||| ||||||||| | ||  ||||
Db   91429 TAATCTTTTCTAATGCTTTGACCAAATTTTGTGTAATTAATA 91470


RESULT   8
```

```
AC025797/c
LOCUS       AC025797   161733 bp    DNA             HTG       25-JUN-2000
DEFINITION  Homo sapiens clone RP11-385C22, WORKING DRAFT SEQUENCE, 32
            unordered pieces.
ACCESSION   AC025797
VERSION     AC025797.3  GI:8705135
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161733)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-385C22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161733)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 25, 2000 this sequence version replaced gi:7534070.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L8166
                Center clone name: 385_C_22
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 139384 bases at least Q40
                Consensus quality: 151678 bases at least Q30
                Consensus quality: 156111 bases at least Q20
                Insert size: 171000; agarose-fp
                Insert size: 158633; sum-of-contigs
                Quality coverage: 3.3 in Q20 bases; agarose-fp
                Quality coverage: 3.6 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 32 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1167: contig of 1167 bp in length
```

```
          *     1168 1267: gap of      100 bp
          *     1268     2620: contig of 1353 bp in length
          *     2621 2720: gap of      100 bp
          *     2721     4325: contig of 1605 bp in length
          *     4326 4425: gap of      100 bp
          *     4426     5495: contig of 1070 bp in length
          *     5496 5595: gap of      100 bp
          *     5596     6690: contig of 1095 bp in length
          *     6691 6790: gap of      100 bp
          *     6791     8804: contig of 2014 bp in length
          *     8805 8904: gap of      100 bp
          *     8905     9390: contig of 486 bp in length
          *     9391 9490: gap of      100 bp
          *     9491    10893: contig of 1403 bp in length
          *    10894 10993: gap of      100 bp
          *    10994    12705: contig of 1712 bp in length
          *    12706 12805: gap of      100 bp
          *    12806    14698: contig of 1893 bp in length
          *    14699 14798: gap of      100 bp
          *    14799    16989: contig of 2191 bp in length
          *    16990 17089: gap of      100 bp
          *    17090    19915: contig of 2826 bp in length
          *    19916 20015: gap of      100 bp
          *    20016    23312: contig of 3297 bp in length
          *    23313 23412: gap of      100 bp
          *    23413    25456: contig of 2044 bp in length
          *    25457 25556: gap of      100 bp
          *    25557    28737: contig of 3181 bp in length
          *    28738 28837: gap of      100 bp
          *    28838    31790: contig of 2953 bp in length
          *    31791 31890: gap of      100 bp
          *    31891    36413: contig of 4523 bp in length
          *    36414 36513: gap of      100 bp
          *    36514    41435: contig of 4922 bp in length
          *    41436 41535: gap of      100 bp
          *    41536    46569: contig of 5034 bp in length
          *    46570 46669: gap of      100 bp
          *    46670    52443: contig of 5774 bp in length
          *    52444 52543: gap of      100 bp
          *    52544    58987: contig of 6444 bp in length
          *    58988 59087: gap of      100 bp
          *    59088    65801: contig of 6714 bp in length
          *    65802 65901: gap of      100 bp
          *    65902    74132: contig of 8231 bp in length
          *    74133 74232: gap of      100 bp
          *    74233    80963: contig of 6731 bp in length
          *    80964 81063: gap of      100 bp
          *    81064    87836: contig of 6773 bp in length
          *    87837 87936: gap of      100 bp
          *    87937    95449: contig of 7513 bp in length
          *    95450 95549: gap of      100 bp
          *    95550   103931: contig of 8382 bp in length
          *   103932 104031: gap of      100 bp
          *   104032   114614: contig of 10583 bp in length
          *   114615 114714: gap of      100 bp
          *   114715   126025: contig of 11311 bp in length
          *   126026 126125: gap of      100 bp
          *   126126   136164: contig of 10039 bp in length
          *   136165 136264: gap of      100 bp
          *   136265   150282: contig of 14018 bp in length
          *   150283 150382: gap of      100 bp
          *   150383   161733: contig of 11351 bp in length.
FEATURES             Location/Qualifiers
     source          1. .161733
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-385C22"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1167
                     /note="assembly_fragment"
     misc_feature    1268. .2620
                     /note="assembly_fragment"
     misc_feature    2721. .4325
                     /note="assembly_fragment"
     misc_feature    4426. .5495
                     /note="assembly_fragment"
     misc_feature    5596. .6690
                     /note="assembly_fragment"
     misc_feature    6791. .8804
                     /note="assembly_fragment"
     misc_feature    8905. .9390
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    9491. .10893
                     /note="assembly_fragment"
     misc_feature    10994. .12705
                     /note="assembly_fragment"
     misc_feature    12806. .14698
                     /note="assembly_fragment"
     misc_feature    14799. .16989
                     /note="assembly_fragment"
     misc_feature    17090. .19915
                     /note="assembly_fragment"
     misc_feature    20016. .23312
                     /note="assembly_fragment"
     misc_feature    23413. .25456
                     /note="assembly_fragment"
     misc_feature    25557. .28737
                     /note="assembly_fragment"
     misc_feature    28838. .31790
                     /note="assembly_fragment"
     misc_feature    31891. .36413
                     /note="assembly_fragment"
     misc_feature    36514. .41435
                     /note="assembly_fragment"
     misc_feature    41536. .46569
                     /note="assembly_fragment"
     misc_feature    46670. .52443
                     /note="assembly_fragment"
     misc_feature    52544. .58987
                     /note="assembly_fragment"
     misc_feature    59088. .65801
                     /note="assembly_fragment"
     misc_feature    65902. .74132
                     /note="assembly_fragment"
     misc_feature    74233. .80963
                     /note="assembly_fragment"
     misc_feature    81064. .87836
                     /note="assembly_fragment"
     misc_feature    87937. .95449
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    95550. .103931
                     /note="assembly_fragment"
     misc_feature    104032. .114614
                     /note="assembly_fragment"

  Query Match             58.1%;  Score 24.4;  DB 70;  Length 161733;
  Best Local Similarity   73.8%;  Pred. No. 4e+02;
  Matches   31;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 tattctttgctcaaaatttgatcaaattttgagcaaagaata 42
           ||||  || | ||||||||| |||||   | || ||||| ||
Db   91470 TATTAATTACACAAAATTTGGTCAAAGCATTAGAAAAGATTA 91429


RESULT   9
AL442067
LOCUS       AL442067   177099 bp    DNA             HTG       06-FEB-2001
DEFINITION  Homo sapiens chromosome 13 clone RP11-128N14, *** SEQUENCING IN
            PROGRESS ***, 8 unordered pieces.
ACCESSION   AL442067
VERSION     AL442067.13  GI:12709921
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 177099)
  AUTHORS   Cobley,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 7, 2001 this sequence version replaced gi:12657191.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA128N14
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 173769 bases at least Q40
            Consensus quality: 175005 bases at least Q30
            Consensus quality: 175758 bases at least Q20
            Insert size: 176399; sum-of-contigs
            Insert size: 151541; 6.6% error; agarose-fp
            Quality coverage: 11.10x in Q20 bases; sum-of-contigs Quality
            coverage: 16.88x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    93940: contig of 93940 bp in length
            *   93941 94040: gap of      100 bp
            *   94041   163510: contig of 69470 bp in length
            *  163511 163610: gap of      100 bp
            *  163611   165813: contig of 2203 bp in length
            *  165814 165913: gap of      100 bp
            *  165914   168160: contig of 2247 bp in length
            *  168161 168260: gap of      100 bp
            *  168261   170441: contig of 2181 bp in length
            *  170442 170541: gap of      100 bp
            *  170542   172613: contig of 2072 bp in length
            *  172614 172713: gap of      100 bp
            *  172714   174763: contig of 2050 bp in length
            *  174764 174863: gap of      100 bp
            *  174864   177099: contig of 2236 bp in length.
FEATURES             Location/Qualifiers
     source          1. .177099
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-128N14"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .93940
                     /note="assembly_fragment:05863
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    94041. .163510
                     /note="assembly_fragment:04987
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:right"
     misc_feature    163611. .165813
                     /note="assembly_fragment:00275"
     misc_feature    165914. .168160
                     /note="assembly_fragment:01244"
     misc_feature    168261. .170441
                     /note="assembly_fragment:04705"
     misc_feature    170542. .172613
                     /note="assembly_fragment:05675"
     misc_feature    172714. .174763
                     /note="assembly_fragment:05726"
     misc_feature    174864. .177099
                     /note="assembly_fragment:05736"
BASE COUNT    53391 a  34238 c  33618 g  55150 t    702 others
ORIGIN


  Query Match             58.1%;  Score 24.4;  DB 81;  Length 177099;
  Best Local Similarity   73.8%;  Pred. No. 4e+02;
  Matches   31;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 tattctttgctcaaaatttgatcaaattttgagcaaagaata 42
           ||||  || | |||||||||| |||||   | || ||||| ||
Db   50375 TATTAATTACACAAAATTTGGTCAAAGCATTAGAAAAGATTA 50416


RESULT  10
AL442067/c
LOCUS       AL442067   177099 bp    DNA             HTG       06-FEB-2001
DEFINITION  Homo sapiens chromosome 13 clone RP11-128N14, *** SEQUENCING IN
            PROGRESS ***, 8 unordered pieces.
ACCESSION   AL442067
VERSION     AL442067.13  GI:12709921
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 177099)
  AUTHORS   Cobley,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 7, 2001 this sequence version replaced gi:12657191.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA128N14
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 173769 bases at least Q40
            Consensus quality: 175005 bases at least Q30
            Consensus quality: 175758 bases at least Q20
            Insert size: 176399; sum-of-contigs
            Insert size: 151541; 6.6% error; agarose-fp
            Quality coverage: 11.10x in Q20 bases; sum-of-contigs Quality
            coverage: 16.88x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    93940: contig of 93940 bp in length
            *   93941 94040: gap of      100 bp
            *   94041   163510: contig of 69470 bp in length
            *  163511 163610: gap of      100 bp
            *  163611   165813: contig of 2203 bp in length
```

```
            *   165814 165913: gap of      100 bp
            *   165914   168160: contig of 2247 bp in length
            *   168161 168260: gap of      100 bp
            *   168261   170441: contig of 2181 bp in length
            *   170442 170541: gap of      100 bp
            *   170542   172613: contig of 2072 bp in length
            *   172614 172713: gap of      100 bp
            *   172714   174763: contig of 2050 bp in length
            *   174764 174863: gap of      100 bp
            *   174864   177099: contig of 2236 bp in length.
FEATURES             Location/Qualifiers
     source          1. .177099
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-128N14"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .93940
                     /note="assembly_fragment:05863
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    94041. .163510
                     /note="assembly_fragment:04987
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:right"
     misc_feature    163611. .165813
                     /note="assembly_fragment:00275"
     misc_feature    165914. .168160
                     /note="assembly_fragment:01244"
     misc_feature    168261. .170441
                     /note="assembly_fragment:04705"
     misc_feature    170542. .172613
                     /note="assembly_fragment:05675"
     misc_feature    172714. .174763
                     /note="assembly_fragment:05726"
     misc_feature    174864. .177099
                     /note="assembly_fragment:05736"
BASE COUNT    53391 a  34238 c  33618 g  55150 t    702 others
ORIGIN


  Query Match             58.1%;  Score 24.4;  DB 81;  Length 177099;
  Best Local Similarity   73.8%;  Pred. No. 4e+02;
  Matches   31;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 tattctttgctcaaaatttgatcaaattttgagcaaagaata 42
           || ||||| || |   ||||| |||||||||| | ||  ||||
Db   50416 TAATCTTTTCTAATGCTTTGACCAAATTTTGTGTAATTAATA 50375


RESULT  11
AC006322
LOCUS       AC006322   179640 bp    DNA             PRI       21-DEC-1999
DEFINITION  Homo sapiens PAC clone RP5-1060B11 from 7q11.23-q21.1, complete
            sequence.
ACCESSION   AC006322
VERSION     AC006322.2  GI:4454515
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179640)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 179640)
  AUTHORS   Holmes,A. and Gregory,S.
  TITLE     The sequence of Homo sapiens PAC clone RP5-1060B11
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 179640)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 179640)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 179640)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Mar 19, 1999 this sequence version replaced gi:4199987.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_DJ1060B11
            --------------

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-5, prepared by
            Pieter de Jong and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu) using the method described by
            Ioannou et al., Nature Genetics 6:84-9 (1994).  The library is from
            one male donor.
            The clone may be obtained either from Genome Systems, Inc.
            (http://www.genomesystems.com) or Research Genetics, Inc.
            (http://www.resgen.com); or from Pieter de Jong.
            VECTOR:  pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            Actual start of this clone is at base position 1 of RP5-1060B11;
            actual end is at 179640 of RP5-1060B11.
FEATURES             Location/Qualifiers
     source          1. .179640
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q11.23-q21.1"
                     /clone="RP5-1060B11"
                     /clone_lib="RPCI-5"
     repeat_region   1. .866
                     /rpt_family="MER2_type"
     repeat_region   990. .1280
```

```
                /rpt_family="Alu"
repeat_region   2419. .2718
                /rpt_family="Alu"
repeat_region   3677. .9682
                /rpt_family="L1"
repeat_region   10069. .10158
                /rpt_family="Alu"
repeat_region   10297. .10335
                /rpt_family="AT_rich"
repeat_region   10530. .10784
                /rpt_family="Alu"
repeat_region   10787. .10811
                /rpt_family="(CAA)n"
repeat_region   11100. .11231
                /rpt_family="L2"
repeat_region   11346. .11448
                /rpt_family="L2"
repeat_region   11862. .11925
                /rpt_family="AT_rich"
repeat_region   11966. .12012
                /rpt_family="purine-rich"
repeat_region   12013. .12048
                /rpt_family="(CA)n"
repeat_region   12049. .12084
                /rpt_family="purine-rich"
repeat_region   12096. .12301
                /rpt_family="MIR"
repeat_region   12783. .13327
                /rpt_family="MER90"
gene            <13542. .37648
                /gene="WUGSC:H_DJ1060B11.1"
CDS             join(<13542. .13583,17541. .17698,21823. .21887,35672. .35814,
                37193. .37648)
                /gene="WUGSC:H_DJ1060B11.1"
                /note="match to D49423 (PID:g631372); H_DJ1060B11.1"
                /codon_start=1
                /product="semaphorin III precursor"
                /protein_id="AAD20933.1"
                /db_xref="GI:4454516"
                /translation="EPTAISAMELSTKQQQLYIGSTAGVAQLPLHRCDIYGKACAECC
                LARDPYCAWDGSACSRYFPTAKRRTRRQDIRNGDPLTHCSDLHHDNHHGHSPEERIIY
                GVENSSTFLECSPKSQRALVYWQFQRRNEERKEEIRVDDHIIRTDQGLLLRSLQQKDS
                GNYLCHAVEHGFIQTLLKVTLEVIDTEHLEELLHKDDDGDGSKTKEMSNSMTPSQKVW
                YRDFMQLINHPNLNTMDEFCEQVWKRDRKQRRQRPGHTPGNSNKWKHLQENKKGRNRR
                THEFERAPRSV"
repeat_region   13955. .13979
                /rpt_family="AT_rich"
repeat_region   14388. .14419
                /rpt_family="POLY_A"
repeat_region   14810. .14832
                /rpt_family="AT_rich"
repeat_region   15750. .15850
                /rpt_family="purine-rich"
repeat_region   16749. .16878
                /rpt_family="MIR"
repeat_region   18296. .18394
                /rpt_family="(TA)n"
repeat_region   18395. .18697
                /rpt_family="Alu"
repeat_region   18698. .18715
                /rpt_family="(TA)n"
repeat_region   19588. .19885
                /rpt_family="Alu"
repeat_region   20918. .21227
                /rpt_family="Alu"
repeat_region   21724. .21746
                /rpt_family="AT_rich"
repeat_region   22057. .22113
                /rpt_family="AT_rich"
repeat_region   26849. .26884
                /rpt_family="AT_rich"
repeat_region   26910. .27029
                /rpt_family="(TA)n"
repeat_region   27573. .27873
                /rpt_family="Alu"
repeat_region   27931. .27978
                /rpt_family="AT_rich"
repeat_region   29914. .30219
                /rpt_family="Alu"
repeat_region   31008. .31224
                /rpt_family="L1"
repeat_region   31246. .31545
                /rpt_family="Alu"
repeat_region   31786. .31806
                /rpt_family="AT_rich"
repeat_region   31992. .32008
                /rpt_family="AT_rich"
repeat_region   32009. .32310
                /rpt_family="Alu"
repeat_region   32311. .32329
                /rpt_family="AT_rich"
repeat_region   32803. .32830
                /rpt_family="AT_rich"
repeat_region   33041. .33350
                /rpt_family="Alu"
repeat_region   33352. .33391
                /rpt_family="(TA)n"
repeat_region   33781. .33912
                /rpt_family="Alu"
repeat_region   34707. .34735
                /rpt_family="AT_rich"
misc_feature    35755. .35814
                /gene="WUGSC:H_DJ1060B11.1"
                /note="match to EST R19784 (NID:g774418) yg28c12.r1"
repeat_region   35942. .36038
                /rpt_family="L2"
repeat_region   36148. .36180
                /rpt_family="AT_rich"

  Query Match             58.1%;  Score 24.4;  DB 86;  Length 179640;
  Best Local Similarity   73.8%;  Pred. No. 4e+02;
  Matches   31;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaagaata 42
          ||| ||| ||| |||||||||  | ||  | |||||| || |
Db 134539 TATACTTAGCTTAAAATTTGAGTATATCATAAGCAAAAAACA 134580


RESULT  12
AC006322/c
LOCUS       AC006322   179640 bp    DNA             PRI       21-DEC-1999
DEFINITION  Homo sapiens PAC clone RP5-1060B11 from 7q11.23-q21.1, complete
            sequence.
ACCESSION   AC006322
VERSION     AC006322.2  GI:4454515
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179640)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 179640)
  AUTHORS   Holmes,A. and Gregory,S.
  TITLE     The sequence of Homo sapiens PAC clone RP5-1060B11
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 179640)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 179640)
```

```
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 179640)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Mar 19, 1999 this sequence version replaced gi:4199987.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_DJ1060B11
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center.  For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            This clone was derived from human PAC library RPCI-5, prepared by
            Pieter de Jong and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu) using the method described by
            Ioannou et al., Nature Genetics 6:84-9 (1994).  The library is from
            one male donor.
            The clone may be obtained either from Genome Systems, Inc.
            (http://www.genomesystems.com) or Research Genetics, Inc.
            (http://www.resgen.com); or from Pieter de Jong.
            VECTOR:  pCYPAC2
            NEIGHBORING SEQUENCE INFORMATION:
            Actual start of this clone is at base position 1 of RP5-1060B11;
            actual end is at 179640 of RP5-1060B11.
FEATURES              Location/Qualifiers
     source           1. .179640
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="7"
                      /map="7q11.23-q21.1"
                      /clone="RP5-1060B11"
                      /clone_lib="RPCI-5"
     repeat_region    1. .866
                      /rpt_family="MER2_type"
     repeat_region    990. .1280
                      /rpt_family="Alu"
     repeat_region    2419. .2718
                      /rpt_family="Alu"
     repeat_region    3677. .9682
                      /rpt_family="L1"
     repeat_region    10069. .10158
                      /rpt_family="Alu"
     repeat_region    10297. .10335
                      /rpt_family="AT_rich"
     repeat_region    10530. .10784
                      /rpt_family="Alu"
     repeat_region    10787. .10811
                      /rpt_family="(CAA)n"
     repeat_region    11100. .11231
                      /rpt_family="L2"
     repeat_region    11346. .11448
                      /rpt_family="L2"
     repeat_region    11862. .11925
                      /rpt_family="AT_rich"
     repeat_region    11966. .12012
                      /rpt_family="purine-rich"
     repeat_region    12013. .12048
                      /rpt_family="(CA)n"
     repeat_region    12049. .12084
                      /rpt_family="purine-rich"
     repeat_region    12096. .12301
                      /rpt_family="MIR"
     repeat_region    12783. .13327
                      /rpt_family="MER90"
     gene             <13542. .37648
                      /gene="WUGSC:H_DJ1060B11.1"
     CDS              join(<13542. .13583,17541. .17698,21823. .21887,35672. .35814,
                      37193. .37648)
                      /gene="WUGSC:H_DJ1060B11.1"
                      /note="match to D49423 (PID:g631372); H_DJ1060B11.1"
                      /codon_start=1
                      /product="semaphorin III precursor"
                      /protein_id="AAD20933.1"
                      /db_xref="GI:4454516"
                      /translation="EPTAISAMELSTKQQQLYIGSTAGVAQLPLHRCDIYGKACAECC
                      LARDPYCAWDGSACSRYFPTAKRRTRRQDIRNGDPLTHCSDLHHDNHHGHSPEERIIY
                      GVENSSTFLECSPKSQRALVYWQFQRRNEERKEEIRVDDHIIRTDQGLLLRSLQQKDS
                      GNYLCHAVEHGFIQTLLKVTLEVIDTEHLEELLHKDDDGDGSKTKEMSNSMTPSQKVW
                      YRDFMQLINHPNLNTMDEFCEQVWKRDRKQRRQRPGHTPGNSNKWKHLQENKKGRNRR
                      THEFERAPRSV"
     repeat_region    13955. .13979
                      /rpt_family="AT_rich"
     repeat_region    14388. .14419
                      /rpt_family="POLY_A"
     repeat_region    14810. .14832
                      /rpt_family="AT_rich"
     repeat_region    15750. .15850
                      /rpt_family="purine-rich"
     repeat_region    16749. .16878
                      /rpt_family="MIR"
     repeat_region    18296. .18394
                      /rpt_family="(TA)n"
     repeat_region    18395. .18697
                      /rpt_family="Alu"
     repeat_region    18698. .18715
                      /rpt_family="(TA)n"
     repeat_region    19588. .19885
                      /rpt_family="Alu"
     repeat_region    20918. .21227
                      /rpt_family="Alu"
     repeat_region    21724. .21746
                      /rpt_family="AT_rich"
     repeat_region    22057. .22113
                      /rpt_family="AT_rich"
     repeat_region    26849. .26884
                      /rpt_family="AT_rich"
     repeat_region    26910. .27029
                      /rpt_family="(TA)n"
     repeat_region    27573. .27873
                      /rpt_family="Alu"
     repeat_region    27931. .27978
                      /rpt_family="AT_rich"
     repeat_region    29914. .30219
                      /rpt_family="Alu"
     repeat_region    31008. .31224
```

```
                     /rpt_family="L1"
     repeat_region   31246. .31545
                     /rpt_family="Alu"
     repeat_region   31786. .31806
                     /rpt_family="AT_rich"
     repeat_region   31992. .32008
                     /rpt_family="AT_rich"
     repeat_region   32009. .32310
                     /rpt_family="Alu"
     repeat_region   32311. .32329
                     /rpt_family="AT_rich"
     repeat_region   32803. .32830
                     /rpt_family="AT_rich"
     repeat_region   33041. .33350
                     /rpt_family="Alu"
     repeat_region   33352. .33391
                     /rpt_family="(TA)n"
     repeat_region   33781. .33912
                     /rpt_family="Alu"
     repeat_region   34707. .34735
                     /rpt_family="AT_rich"
     misc_feature    35755. .35814
                     /gene="WUGSC:H_DJ1060B11.1"
                     /note="match to EST R19784 (NID:g774418) yg28c12.r1"
     repeat_region   35942. .36038
                     /rpt_family="L2"
     repeat_region   36148. .36180
                     /rpt_family="AT_rich"

  Query Match             58.1%;  Score 24.4;  DB 86;  Length 179640;
  Best Local Similarity   73.8%;  Pred. No. 4e+02;
  Matches   31;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaagaata 42
          | || |||||| |  || |  ||||||||| ||| ||| |||
Db 134580 TGTTTTTTGCTTATGATATACTCAAATTTTAAGCTAAGTATA 134539


RESULT  13
AC073498
LOCUS       AC073498   183285 bp    DNA             HTG       27-OCT-2000
DEFINITION  Homo sapiens chromosome 7 clone RP11-242J14, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC073498
VERSION     AC073498.2  GI:9211373
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 183285)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 183285)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jul 15, 2000 this sequence version replaced gi:8572511.

            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0242J14
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 97%
            Sequencing vector: plasmid; 3%
            Chemistry: Dye-primer ET; 97% of reads
            Chemistry: Dye-terminator Big Dye; 3% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 180801 bases at least Q40
            Consensus quality: 181757 bases at least Q30
            Consensus quality: 182111 bases at least Q20
            Insert size: 193000; agarose-fp
            Insert size: 182885; sum-of-contigs
            Quality coverage: 5.08 in Q20 bases; agarose-fp
            Quality coverage: 5.54 in Q20 bases; sum-of-contigs
            ----------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1     6773: contig of 6773 bp in length
            *      6774     6873: gap of unknown length
            *      6874    26451: contig of 19578 bp in length
            *     26452    26551: gap of unknown length
            *     26552    66536: contig of 39985 bp in length
            *     66537    66636: gap of unknown length
            *     66637   110478: contig of 43842 bp in length
            *    110479   110578: gap of unknown length
            *    110579   183285: contig of 72707 bp in length.
FEATURES             Location/Qualifiers
     source          1. .183285
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="RP11-242J14"
     misc_feature    1. .6773
                     /note="assembly_name:Contig4"
     misc_feature    6874. .26451
                     /note="assembly_name:Contig5"
     misc_feature    26552. .66536
                     /note="assembly_name:Contig6
                     clone_end:SP6
                     vector_side:right"
     misc_feature    66637. .110478
                     /note="assembly_name:Contig7"
     misc_feature    110579. .183285
                     /note="assembly_name:Contig8"
BASE COUNT    57905 a  31183 c  32413 g  61384 t    400 others
ORIGIN


  Query Match             58.1%;  Score 24.4;  DB 75;  Length 183285;
  Best Local Similarity   73.8%;  Pred. No. 3.9e+02;
  Matches   31;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaagaata 42
          | || |||||| |  || |  ||||||||| ||| ||| |||
Db  94749 TGTTTTTTGCTTATGATATACTCAAATTTTAAGCTAAGTATA 94790


RESULT  14
AC073498/c
LOCUS       AC073498   183285 bp    DNA             HTG       27-OCT-2000
DEFINITION  Homo sapiens chromosome 7 clone RP11-242J14, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC073498
VERSION     AC073498.2  GI:9211373
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 183285)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 183285)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jul 15, 2000 this sequence version replaced gi:8572511.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0242J14
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 97%
            Sequencing vector: plasmid; 3%
            Chemistry: Dye-primer ET; 97% of reads
            Chemistry: Dye-terminator Big Dye; 3% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 180801 bases at least Q40
            Consensus quality: 181757 bases at least Q30
            Consensus quality: 182111 bases at least Q20
            Insert size: 193000; agarose-fp
            Insert size: 182885; sum-of-contigs
            Quality coverage: 5.08 in Q20 bases; agarose-fp
            Quality coverage: 5.54 in Q20 bases; sum-of-contigs
            ----------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     6773: contig of 6773 bp in length
            *    6774     6873: gap of unknown length
            *    6874    26451: contig of 19578 bp in length
            *   26452    26551: gap of unknown length
            *   26552    66536: contig of 39985 bp in length
            *   66537    66636: gap of unknown length
            *   66637   110478: contig of 43842 bp in length
            *  110479   110578: gap of unknown length
            *  110579   183285: contig of 72707 bp in length.
FEATURES             Location/Qualifiers
     source          1. .183285
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="RP11-242J14"
     misc_feature    1. .6773
                     /note="assembly_name:Contig4"
     misc_feature    6874. .26451
                     /note="assembly_name:Contig5"
     misc_feature    26552. .66536
                     /note="assembly_name:Contig6
                     clone_end:SP6
                     vector_side:right"
     misc_feature    66637. .110478
                     /note="assembly_name:Contig7"
     misc_feature    110579. .183285
                     /note="assembly_name:Contig8"
BASE COUNT    57905 a  31183 c  32413 g  61384 t    400 others
ORIGIN


  Query Match           58.1%;  Score 24.4;  DB 75;  Length 183285;
  Best Local Similarity  73.8%;  Pred. No. 3.9e+02;
  Matches   31;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaagaata 42
          ||| ||| ||| |||||||||  | ||  | |||||| || |

Db  94790 TATACTTAGCTTAAAATTTGAGTATATCATAAGCAAAAAACA 94749


RESULT  15
AC073609
LOCUS       AC073609   212307 bp    DNA              HTG       08-JAN-2001
DEFINITION  Mus musculus clone RP3-518O10, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   AC073609
VERSION     AC073609.17  GI:12043813
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 212307)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
            and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 212307)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jan 7, 2001 this sequence version replaced gi:12000241.
            ---------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
                Center project name: HBFY
                Center clone name: RP3-518O10
            ---------- Summary Statistics
```

```
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 12% of reads
                Chemistry: Dye-terminator Big Dye: 88% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 206544 bases at least Q40
                Consensus quality: 209406 bases at least Q30
                Consensus quality: 211582 bases at least Q20
                Estimated insert size: 209424; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 6.7x in Q20 bases; sum-of-contigs estimation
              --------------
              * NOTE: Estimated insert size may differ from sequence length
              *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 7 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *        1    69093: contig of 69093 bp in length
              *    69094    69193: gap of unknown length
              *    69194   122186: contig of 52993 bp in length
              *   122187   122286: gap of unknown length
              *   122287   155591: contig of 33305 bp in length
              *   155592   155691: gap of unknown length
              *   155692   188568: contig of 32877 bp in length
              *   188569   188668: gap of unknown length
              *   188669   204180: contig of 15512 bp in length
              *   204181   204280: gap of unknown length
              *   204281   211158: contig of 6878 bp in length
              *   211159   211258: gap of unknown length
              *   211259   212307: contig of 1049 bp in length.
FEATURES              Location/Qualifiers
     source           1. .212307
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone="RP3-518O10"
BASE COUNT     67653 a  39685 c  38593 g  64782 t   1594 others
ORIGIN


  Query Match             58.1%;  Score 24.4;  DB 75;  Length 212307;
  Best Local Similarity   82.4%;  Pred. No. 3.9e+02;
  Matches   28;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       2 attctttgctcaaaatttgatcaaattttgagca 35
           |||||||||  | |||||||||  |||| ||||
Db 205765 ATTCTTTGCTATATATTTGATCATCTTTTAAGCA 205798


RESULT  16
AC073609/c
LOCUS       AC073609    212307 bp    DNA             HTG       08-JAN-2001
DEFINITION  Mus musculus clone RP3-518O10, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   AC073609
VERSION     AC073609.17  GI:12043813
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 212307)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
            and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 212307)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jan 7, 2001 this sequence version replaced gi:12000241.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HBFY
                Center clone name: RP3-518O10
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 12% of reads
                Chemistry: Dye-terminator Big Dye: 88% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 206544 bases at least Q40
                Consensus quality: 209406 bases at least Q30
                Consensus quality: 211582 bases at least Q20
                Estimated insert size: 209424; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 6.7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    69093: contig of 69093 bp in length
            *    69094    69193: gap of unknown length
            *    69194   122186: contig of 52993 bp in length
```

```
            *   122187   122286: gap of unknown length
            *   122287   155591: contig of 33305 bp in length
            *   155592   155691: gap of unknown length
            *   155692   188568: contig of 32877 bp in length
            *   188569   188668: gap of unknown length
            *   188669   204180: contig of 15512 bp in length
            *   204181   204280: gap of unknown length
            *   204281   211158: contig of 6878 bp in length
            *   211159   211258: gap of unknown length
            *   211259   212307: contig of 1049 bp in length.
FEATURES             Location/Qualifiers
     source          1. .212307
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP3-518O10"
BASE COUNT    67653 a  39685 c  38593 g  64782 t   1594 others
ORIGIN


  Query Match             58.1%;  Score 24.4;  DB 75;  Length 212307;
  Best Local Similarity   82.4%;  Pred. No. 3.9e+02;
  Matches   28;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       8 tgctcaaaatttgatcaaattttgagcaaagaat 41
           |||| ||||  ||||||||| |  ||||||||||
Db 205798 TGCTTAAAAGATGATCAAATATATAGCAAAGAAT 205765


RESULT  17
HSJ718P11
LOCUS       HSJ718P11  104913 bp    DNA             PRI       15-NOV-2000
DEFINITION  Human DNA sequence from clone RP4-718P11 on chromosome 20p12.1-12.3
            Contains 2 isoforms for part of the gene for a novel class II
            aminotransferase similar to serine palmotyltransferase, ESTs, STSs
            and GSSs, complete sequence.
ACCESSION   AL109983
VERSION     AL109983.2  GI:9368492
KEYWORDS    HTG; aminotransferase; serine palmotyltransferase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 104913)
  AUTHORS   Matthews,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 22, 2000 this sequence version replaced gi:5777578.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP4-718P11 is
            from the library RPCI-4 constructed at the Roswell Park Cancer
            Institute by the group of Pieter de Jong. For further details see
            http://bacpac.med.buffalo.edu/
            VECTOR: pCYPAC2
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            This sequence is the entire insert of clone RP4-718P11 The true
            left end of clone RP5-1077I2 is at 71329 in this sequence. The true
            right end of clone RP11-303K20 is at 39954 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .104913
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="p12.1-12.3"
                     /clone="RP4-718P11"
                     /clone_lib="RPCI-4"
     repeat_region   10. .329
                     /note="MER87 repeat: matches 211. .540 of consensus"
     repeat_region   544. .611
                     /note="MER5A repeat: matches 121. .189 of consensus"
     repeat_region   824. .852
                     /note="MER5B repeat: matches 150. .178 of consensus"
     repeat_region   865. .1291
                     /note="L2 repeat: matches 307. .860 of consensus"
     repeat_region   1565. .1604
                     /note="10 copies 4 mer ttat 100% conserved"
     repeat_region   1643. .2345
                     /note="L1PB3 repeat: matches 5450. .6150 of consensus"
     repeat_region   2534. .2687
                     /note="L2 repeat: matches 833. .1002 of consensus"
     repeat_region   2704. .3333
                     /note="L2 repeat: matches 1774. .2440 of consensus"
     repeat_region   3898. .4038
                     /note="L1MB7 repeat: matches 5776. .5920 of consensus"
     repeat_region   4108. .4463
                     /note="L1MB7 repeat: matches 5303. .5661 of consensus"
     repeat_region   4829. .5139
                     /note="AluSx repeat: matches 4. .312 of consensus"
     repeat_region   5208. .5388
                     /note="AluSg/x repeat: matches 127. .309 of consensus"
     repeat_region   5435. .5486
                     /note="26 copies 2 mer tt 84% conserved"
     repeat_region   5489. .5786
                     /note="AluSx repeat: matches 1. .309 of consensus"
     repeat_region   6844. .6959
                     /note="58 copies 2 mer tt 62% conserved"
     repeat_region   6978. .7364
                     /note="L1MA9 repeat: matches 5894. .6275 of consensus"
     repeat_region   7354. .7682
                     /note="L1 repeat: matches 4721. .5050 of consensus"
     repeat_region   7917. .8004
                     /note="22 copies 4 mer tttc 63% conserved"
     repeat_region   8227. .8292
                     /note="33 copies 2 mer ta 95% conserved"
     repeat_region   8229. .8292
                     /note="16 copies 4 mer tata 96% conserved"
     repeat_region   10536. .10888
                     /note="MER1B repeat: matches 1. .337 of consensus"
     repeat_region   11927. .12045
                     /note="L2 repeat: matches 2143. .2290 of consensus"
     repeat_region   12303. .12811
                     /note="MER75 repeat: matches 6. .514 of consensus"
     gene            13443. .101318
                     /gene="dJ718P11.1"
     CDS             join(<13443. .13597,15536. .15684,20936. .>21054)
                     /gene="dJ718P11.1"
                     /note="match: ESTs: Em:AA192501"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="dJ718P11.1.2 (novel class II aminotransferase
                     similar to serine palmotyltransferase (isoform 2))"
                     /protein_id="CAB65866.1"
                     /db_xref="GI:6691981"
                     /db_xref="SPTREMBL:Q9UGB5"
                     /translation="DFVPLYQDFENFYTRNLYMRIRDNWNRPICSAPGPLFDLMERVS
                     DDYNWTFRFTGRVIKDVINMGSYNFLGLAAKYDESMRTIKDVLEVYGTGVASTRHEMG
                     EFTSSWQNALCVGGVLQADSVPQGPPWSWKMGRHSSLLE"
```

```
CDS             join(<13443. .13597,15536. .15684,32270. .32394,34670. .34763,
                51298. .51403,58692. .58911,67777. .67903,95189. .95324,
                101189. .>101318)
                /gene="dJ718P11.1"
                /note="match: cDNAs: Em:AB011098 Em:Y08686 Em:X95642
                Em:AF004830 Em:U27455 Em:U15555
                match: ESTs: Em:AA357446 Em:AA192501
                match: proteins: Tr:O54694 Tr:Q9XVI6 Tr:O15270 Tr:P97363
                Tr:Q20375 Tr:Q9XGB0 Sw:Q09925 Sw:P40970 Sw:P48241"
                /codon_start=1
                /evidence=not_experimental
                /product="dJ718P11.1.1 (novel class II aminotransferase
                similar to serine palmotyltransferase (isoform 1))"
                /protein_id="CAB65865.1"
                /db_xref="GI:6691980"
                /db_xref="SPTREMBL:Q9UGB6"
                /translation="DFVPLYQDFENFYTRNLYMRIRDNWNRPICSAPGPLFDLMERVS
                DDYNWTFRFTGRVIKDVINMGSYNFLGLAAKYDESMRTIKDVLEVYGTGVASTRHEMG
                TLDKHKELEDLVAKFLNVEAAMVFGMGFATNSMNIPALVGKGCLILSDELNHTSLVLG
                ARLSGATIRIFKHNNTQSLEKLLRDAVIYGQPRTRRAWKKILILVEGVYSMEGSIVHL
                PQIIALKKKYKAYLYIDEAHSIGAVGPTGRGVTEFFGLDPHEVDVLMGTFTKSFGASG
                GYIAGRKDLVDYLRVHSHSAVYASSMSPPIAEQIIRSLKLIMGLDGTTQGLQRVQQLA
                KNTRYFRQRLQEMGFIIYGNENASVVPLLLYMPGKVAAFARHMLEKKIGVVVVGFPAT
                PLAEARARFCVSAAHTREMLDT"
repeat_region   14113. .14582
                /note="L1MA4 repeat: matches 5815. .6292 of consensus"
repeat_region   15902. .16080
                /note="L2 repeat: matches 2480. .2658 of consensus"
misc_feature    complement(16079. .16250)
                /note="match: STS: Em:HSPF04H5"
repeat_region   18972. .19148
                /note="AluJo repeat: matches 121. .294 of consensus"
misc_feature    19894. .20269
                /gene="dJ718P11.1"
                /note="match: GSS: Em:AQ044246"
repeat_region   20181. .20417
                /note="L1M4 repeat: matches 4366. .4597 of consensus"
misc_feature    complement(20538. .20941)
                /note="match: GSS: Em:AQ662868"
repeat_region   21219. .21378
                /note="L1M4 repeat: matches 5160. .5326 of consensus"
repeat_region   21415. .21708
                /note="AluSc repeat: matches 1. .290 of consensus"
repeat_region   22230. .22469
                /note="LTR23 repeat: matches 206. .437 of consensus"
repeat_region   23617. .23863
                /note="MER58B repeat: matches 79. .337 of consensus"
repeat_region   24767. .24827
                /note="MIR repeat: matches 63. .123 of consensus"
repeat_region   25619. .25838
                /note="MIR repeat: matches 31. .238 of consensus"
repeat_region   25845. .26034
                /note="MIR repeat: matches 66. .258 of consensus"
misc_feature    complement(25953. .26556)
                /note="match: GSS: Em:AQ482745"
repeat_region   27967. .28282
                /note="AluJo repeat: matches 1. .312 of consensus"
repeat_region   28751. .28859
                /note="L2 repeat: matches 2620. .2723 of consensus"
repeat_region   29512. .29809
                /note="AluY repeat: matches 1. .309 of consensus"
repeat_region   29810. .30116
                /note="AluSx repeat: matches 1. .296 of consensus"
repeat_region   30520. .30690
                /note="Charlie5 repeat: matches 2414. .2584 of consensus"
repeat_region   30691. .30996
                /note="AluSx repeat: matches 1. .306 of consensus"
repeat_region   30997. .31098
                /note="Charlie5 repeat: matches 2313. .2414 of consensus"
repeat_region   31097. .31278
                /note="Charlie5 repeat: matches -5. .115 of consensus"
repeat_region   31450. .31581
                /note="MIR repeat: matches 105. .249 of consensus"
misc_feature    complement(31474. .32053)
                /note="match: GSS: Em:AQ083543"
repeat_region   31753. .32181
                /note="L1ME3 repeat: matches 5679. .6131 of consensus"
misc_feature    32059. .32525
                /gene="dJ718P11.1"
                /note="match: GSS: Em:AQ774405"
misc_feature    32081. .32543
                /gene="dJ718P11.1"
                /note="match: GSS: Em:AQ630217"
repeat_region   32484. .32785

  Query Match            57.6%;  Score 24.2;  DB 92;  Length 104913;
  Best Local Similarity  78.4%;  Pred. No. 4.6e+02;
  Matches   29;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaa 37
          ||| |||||||||| ||||| || ||| |||  | |||
Db   4431 TATCCTTTGCTCATAATTTTATTAAACTTTTTGAAAA 4467


RESULT  18
HSJ718P11/c
LOCUS       HSJ718P11  104913 bp    DNA             PRI       15-NOV-2000
DEFINITION  Human DNA sequence from clone RP4-718P11 on chromosome 20p12.1-12.3
            Contains 2 isoforms for part of the gene for a novel class II
            aminotransferase similar to serine palmotyltransferase, ESTs, STSs
            and GSSs, complete sequence.
ACCESSION   AL109983
VERSION     AL109983.2  GI:9368492
KEYWORDS    HTG; aminotransferase; serine palmotyltransferase.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 104913)
  AUTHORS   Matthews,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 22, 2000 this sequence version replaced gi:5777578.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP4-718P11 is
            from the library RPCI-4 constructed at the Roswell Park Cancer
            Institute by the group of Pieter de Jong. For further details see
            http://bacpac.med.buffalo.edu/
            VECTOR: pCYPAC2
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            This sequence is the entire insert of clone RP4-718P11 The true
            left end of clone RP5-1077I2 is at 71329 in this sequence. The true
            right end of clone RP11-303K20 is at 39954 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .104913
                     /organism="Homo sapiens"
```

```
                /db_xref="taxon:9606"
                /chromosome="20"
                /map="p12.1-12.3"
                /clone="RP4-718P11"
                /clone_lib="RPCI-4"
repeat_region   10. .329
                /note="MER87 repeat: matches 211. .540 of consensus"
repeat_region   544. .611
                /note="MER5A repeat: matches 121. .189 of consensus"
repeat_region   824. .852
                /note="MER5B repeat: matches 150. .178 of consensus"
repeat_region   865. .1291
                /note="L2 repeat: matches 307. .860 of consensus"
repeat_region   1565. .1604
                /note="10 copies 4 mer ttat 100% conserved"
repeat_region   1643. .2345
                /note="L1PB3 repeat: matches 5450. .6150 of consensus"
repeat_region   2534. .2687
                /note="L2 repeat: matches 833. .1002 of consensus"
repeat_region   2704. .3333
                /note="L2 repeat: matches 1774. .2440 of consensus"
repeat_region   3898. .4038
                /note="L1MB7 repeat: matches 5776. .5920 of consensus"
repeat_region   4108. .4463
                /note="L1MB7 repeat: matches 5303. .5661 of consensus"
repeat_region   4829. .5139
                /note="AluSx repeat: matches 4. .312 of consensus"
repeat_region   5208. .5388
                /note="AluSg/x repeat: matches 127. .309 of consensus"
repeat_region   5435. .5486
                /note="26 copies 2 mer tt 84% conserved"
repeat_region   5489. .5786
                /note="AluSx repeat: matches 1. .309 of consensus"
repeat_region   6844. .6959
                /note="58 copies 2 mer tt 62% conserved"
repeat_region   6978. .7364
                /note="L1MA9 repeat: matches 5894. .6275 of consensus"
repeat_region   7354. .7682
                /note="L1 repeat: matches 4721. .5050 of consensus"
repeat_region   7917. .8004
                /note="22 copies 4 mer tttc 63% conserved"
repeat_region   8227. .8292
                /note="33 copies 2 mer ta 95% conserved"
repeat_region   8229. .8292
                /note="16 copies 4 mer tata 96% conserved"
repeat_region   10536. .10888
                /note="MER1B repeat: matches 1. .337 of consensus"
repeat_region   11927. .12045
                /note="L2 repeat: matches 2143. .2290 of consensus"
repeat_region   12303. .12811
                /note="MER75 repeat: matches 6. .514 of consensus"
gene            13443. .101318
                /gene="dJ718P11.1"
CDS             join(<13443. .13597,15536. .15684,20936. .>21054)
                /gene="dJ718P11.1"
                /note="match: ESTs: Em:AA192501"
                /codon_start=1
                /evidence=not_experimental
                /product="dJ718P11.1.2 (novel class II aminotransferase
                similar to serine palmotyltransferase (isoform 2))"
                /protein_id="CAB65866.1"
                /db_xref="GI:6691981"
                /db_xref="SPTREMBL:Q9UGB5"
                /translation="DFVPLYQDFENFYTRNLYMRIRDNWNRPICSAPGPLFDLMERVS
                DDYNWTFRFTGRVIKDVINMGSYNFLGLAAKYDESMRTIKDVLEVYGTGVASTRHEMG
                EFTSSWQNALCVGGVLQADSVPQGPPWSWKMGRHSSLLE"
CDS             join(<13443. .13597,15536. .15684,32270. .32394,34670. .34763,
                51298. .51403,58692. .58911,67777. .67903,95189. .95324,
                101189. .>101318)
                /gene="dJ718P11.1"
                /note="match: cDNAs: Em:AB011098 Em:Y08686 Em:X95642
                Em:AF004830 Em:U27455 Em:U15555
                match: ESTs: Em:AA357446 Em:AA192501
                match: proteins: Tr:O54694 Tr:Q9XVI6 Tr:O15270 Tr:P97363
                Tr:Q20375 Tr:Q9XGB0 Sw:Q09925 Sw:P40970 Sw:P48241"
                /codon_start=1
                /evidence=not_experimental
                /product="dJ718P11.1.1 (novel class II aminotransferase
                similar to serine palmotyltransferase (isoform 1))"
                /protein_id="CAB65865.1"
                /db_xref="GI:6691980"
                /db_xref="SPTREMBL:Q9UGB6"
                /translation="DFVPLYQDFENFYTRNLYMRIRDNWNRPICSAPGPLFDLMERVS
                DDYNWTFRFTGRVIKDVINMGSYNFLGLAAKYDESMRTIKDVLEVYGTGVASTRHEMG
                TLDKHKELEDLVAKFLNVEAAMVFGMGFATNSMNIPALVGKGCLILSDELNHTSLVLG
                ARLSGATIRIFKHNNTQSLEKLLRDAVIYGQPRTRRAWKKILILVEGVYSMEGSIVHL
                PQIIALKKKYKAYLYIDEAHSIGAVGPTGRGVTEFFGLDPHEVDVLMGTFTKSFGASG
                GYIAGRKDLVDYLRVHSHSAVYASSMSPPIAEQIIRSLKLIMGLDGTTQGLQRVQQLA
                KNTRYFRQRLQEMGFIIYGNENASVVPLLLYMPGKVAAFARHMLEKKIGVVVVGFPAT
                PLAEARARFCVSAAHTREMLDT"
repeat_region   14113. .14582
                /note="L1MA4 repeat: matches 5815. .6292 of consensus"
repeat_region   15902. .16080
                /note="L2 repeat: matches 2480. .2658 of consensus"
misc_feature    complement(16079. .16250)
                /note="match: STS: Em:HSPF04H5"
repeat_region   18972. .19148
                /note="AluJo repeat: matches 121. .294 of consensus"
misc_feature    19894. .20269
                /gene="dJ718P11.1"
                /note="match: GSS: Em:AQ044246"
repeat_region   20181. .20417
                /note="L1M4 repeat: matches 4366. .4597 of consensus"
misc_feature    complement(20538. .20941)
                /note="match: GSS: Em:AQ662868"
repeat_region   21219. .21378
                /note="L1M4 repeat: matches 5160. .5326 of consensus"
repeat_region   21415. .21708
                /note="AluSc repeat: matches 1. .290 of consensus"
repeat_region   22230. .22469
                /note="LTR23 repeat: matches 206. .437 of consensus"
repeat_region   23617. .23863
                /note="MER58B repeat: matches 79. .337 of consensus"
repeat_region   24767. .24827
                /note="MIR repeat: matches 63. .123 of consensus"
repeat_region   25619. .25838
                /note="MIR repeat: matches 31. .238 of consensus"
repeat_region   25845. .26034
                /note="MIR repeat: matches 66. .258 of consensus"
misc_feature    complement(25953. .26556)
                /note="match: GSS: Em:AQ482745"
repeat_region   27967. .28282
                /note="AluJo repeat: matches 1. .312 of consensus"
repeat_region   28751. .28859
                /note="L2 repeat: matches 2620. .2723 of consensus"
repeat_region   29512. .29809
                /note="AluY repeat: matches 1. .309 of consensus"
repeat_region   29810. .30116
                /note="AluSx repeat: matches 1. .296 of consensus"
repeat_region   30520. .30690
                /note="Charlie5 repeat: matches 2414. .2584 of consensus"
repeat_region   30691. .30996
                /note="AluSx repeat: matches 1. .306 of consensus"
repeat_region   30997. .31098
                /note="Charlie5 repeat: matches 2313. .2414 of consensus"
repeat_region   31097. .31278
                /note="Charlie5 repeat: matches -5. .115 of consensus"
repeat_region   31450. .31581
                /note="MIR repeat: matches 105. .249 of consensus"
misc_feature    complement(31474. .32053)
                /note="match: GSS: Em:AQ083543"
repeat_region   31753. .32181
                /note="L1ME3 repeat: matches 5679. .6131 of consensus"
misc_feature    32059. .32525
                /gene="dJ718P11.1"
                /note="match: GSS: Em:AQ774405"
```

```
     misc_feature    32081. .32543
                     /gene="dJ718P11.1"
                     /note="match: GSS: Em:AQ630217"
     repeat_region   32484. .32785

  Query Match             57.6%;  Score 24.2;  DB 92;  Length 104913;
  Best Local Similarity   78.4%;  Pred. No. 4.6e+02;
  Matches   29;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      6 tttgctcaaaatttgatcaaattttgagcaaagaata 42
          ||| |  ||| ||| || ||||| |||||||||| |||
Db   4467 TTTTCAAAAAGTTTAATAAAATTATGAGCAAAGGATA 4431


RESULT 19
AC078977
LOCUS       AC078977   126413 bp    DNA             HTG       14-AUG-2000
DEFINITION  Oryza sativa chromosome 5 clone P0496H07, *** SEQUENCING IN
            PROGRESS ***, 10 unordered pieces.
ACCESSION   AC078977
VERSION     AC078977.1  GI:9800546
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
            Oryza.
REFERENCE   1  (bases 1 to 126413)
  AUTHORS   Hsing,Y.-I.C., Chow,T.-Y., Chen,C.-S., Wu,H.-P., Chao,Y.-T. and
            Liu,S.-M.
  TITLE     Oryza sativa PAC P0496H07 genomics sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 126413)
  AUTHORS   Hsing,Y.-I.C. and Chow,T.-Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-2000) Institute of Botany, Academia Sinica, 128,
            Section 2, Yien-chu-yuan Road, Nankang, Taipei 11529, Taiwan
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2807: contig of 2807 bp in length
            *                    gap of unknown length
            *     2808     6593: contig of 3786 bp in length
            *                    gap of unknown length
            *     6594    10791: contig of 4198 bp in length
            *                    gap of unknown length
            *    10792    20096: contig of 9305 bp in length
            *                    gap of unknown length
            *    20097    33022: contig of 12926 bp in length
            *                    gap of unknown length
            *    33023    46251: contig of 13229 bp in length
            *                    gap of unknown length
            *    46252    58018: contig of 11767 bp in length
            *                    gap of unknown length
            *    58019    69273: contig of 11255 bp in length
            *                    gap of unknown length
            *    69274    91013: contig of 21740 bp in length
            *                    gap of unknown length
            *    91014   126413: contig of 35400 bp in length.
FEATURES            Location/Qualifiers
     source         1. .126413
                    /organism="Oryza sativa"
                    /db_xref="taxon:4530"
                    /chromosome="5"
                    /clone="P0496H07"
BASE COUNT    35329 a  27581 c  27793 g  35701 t      9 others
ORIGIN


  Query Match             57.6%;  Score 24.2;  DB 76;  Length 126413;
  Best Local Similarity   78.4%;  Pred. No. 4.6e+02;
  Matches   29;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaa 37
          |||  ||| ||  ||||||||||||||||  || |||
Db  97635 TATGTTTTCCTATAAATTTGATCAAATTTAAAGAAAA 97671


RESULT 20
AC078977/c
LOCUS       AC078977   126413 bp    DNA             HTG       14-AUG-2000
DEFINITION  Oryza sativa chromosome 5 clone P0496H07, *** SEQUENCING IN
            PROGRESS ***, 10 unordered pieces.
ACCESSION   AC078977
VERSION     AC078977.1  GI:9800546
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
            Oryza.
REFERENCE   1  (bases 1 to 126413)
  AUTHORS   Hsing,Y.-I.C., Chow,T.-Y., Chen,C.-S., Wu,H.-P., Chao,Y.-T. and
            Liu,S.-M.
  TITLE     Oryza sativa PAC P0496H07 genomics sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 126413)
  AUTHORS   Hsing,Y.-I.C. and Chow,T.-Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-2000) Institute of Botany, Academia Sinica, 128,
            Section 2, Yien-chu-yuan Road, Nankang, Taipei 11529, Taiwan
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2807: contig of 2807 bp in length
            *                    gap of unknown length
            *     2808     6593: contig of 3786 bp in length
            *                    gap of unknown length
            *     6594    10791: contig of 4198 bp in length
            *                    gap of unknown length
            *    10792    20096: contig of 9305 bp in length
            *                    gap of unknown length
            *    20097    33022: contig of 12926 bp in length
            *                    gap of unknown length
            *    33023    46251: contig of 13229 bp in length
            *                    gap of unknown length
            *    46252    58018: contig of 11767 bp in length
            *                    gap of unknown length
            *    58019    69273: contig of 11255 bp in length
            *                    gap of unknown length
            *    69274    91013: contig of 21740 bp in length
            *                    gap of unknown length
            *    91014   126413: contig of 35400 bp in length.
FEATURES            Location/Qualifiers
     source         1. .126413
                    /organism="Oryza sativa"
                    /db_xref="taxon:4530"
                    /chromosome="5"
                    /clone="P0496H07"
BASE COUNT    35329 a  27581 c  27793 g  35701 t      9 others
ORIGIN


  Query Match             57.6%;  Score 24.2;  DB 76;  Length 126413;
  Best Local Similarity   78.4%;  Pred. No. 4.6e+02;
  Matches   29;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;
```

```
Qy       6 tttgctcaaaatttgatcaaattttgagcaaagaata 42
             ||| ||  ||||||||||||||||  || |||  |||
Db   97671 TTTTCTTTAAATTTGATCAAATTTATAGGAAAACATA 97635
```

Search completed: June 7, 2001, 18:05:31
Job time: 6324 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 7, 2001, 18:09:55 ; Search time 200.8 Seconds
(without alignments)
122.106 Million cell updates/sec

Title: US-09-601-561-2
Perfect score: 42
Sequence: 1 tattctttgctcaaaatttg.....caaattttgagcaaagaata 42

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : N_Geneseq_0401:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	24	57.1	1523	21	A51069	Human LPAAT-gamma-
c	2	24	57.1	1523	21	A51069	Human LPAAT-gamma-
	3	24	57.1	1660	21	A51068	Human LPAAT-gamma-
c	4	24	57.1	1660	21	A51068	Human LPAAT-gamma-
	5	22	52.4	2720	20	Z33573	Human breast tumou
c	6	22	52.4	2720	20	Z33573	Human breast tumou
	7	22	52.4	2975	20	X03792	Human semaphorin E
c	8	22	52.4	2975	20	X03792	Human semaphorin E
	9	22	52.4	2975	21	C66800	Clone BR533_4 codi
c	10	22	52.4	2975	21	C66800	Clone BR533_4 codi
	11	21.6	51.4	715	19	V68977	DNA molecule encod
c	12	21.6	51.4	715	19	V68977	DNA molecule encod
	13	21.6	51.4	715	21	C80976	Human breast tumou
c	14	21.6	51.4	715	21	C80976	Human breast tumou
	15	21.4	51.0	1086	19	V59767	Human secreted pro
c	16	21.4	51.0	1086	19	V59767	Human secreted pro
	17	21.4	51.0	1256	19	V59618	Human secreted pro
c	18	21.4	51.0	1256	19	V59618	Human secreted pro
	19	21.4	51.0	1455	20	Z35927	Streptococcus pneu
c	20	21.4	51.0	1455	20	Z35927	Streptococcus pneu
	21	21.4	51.0	1455	20	Z35933	Streptococcus pneu
c	22	21.4	51.0	1455	20	Z35933	Streptococcus pneu
	23	21.4	51.0	2594	16	T02367	Melon aconitase cD
c	24	21.4	51.0	2594	16	T02367	Melon aconitase cD
	25	21.4	51.0	2738	21	C47165	Arabidopsis thalia
c	26	21.4	51.0	2738	21	C47165	Arabidopsis thalia
	27	21.2	50.5	547	21	C93808	Cat flea hindgut a
c	28	21.2	50.5	547	21	C93808	Cat flea hindgut a
	29	21.2	50.5	600	21	A75883	DNA encoding a 9 k
c	30	21.2	50.5	600	21	A75883	DNA encoding a 9 k
	31	21	50.0	2375	20	Z23128	Rice sucrose trans
c	32	21	50.0	2375	20	Z23128	Rice sucrose trans
	33	20.8	49.5	598	21	C49971	Arabidopsis thalia
c	34	20.8	49.5	598	21	C49971	Arabidopsis thalia
	35	20.8	49.5	681	21	C49961	Arabidopsis thalia
c	36	20.8	49.5	681	21	C49961	Arabidopsis thalia
	37	20.8	49.5	1179	21	F12867	Aspergillus oryzae
c	38	20.8	49.5	1179	21	F12867	Aspergillus oryzae
	39	20.8	49.5	1298	19	V32590	Schwannomin-bindin
c	40	20.8	49.5	1298	19	V32590	Schwannomin-bindin
	41	20.8	49.5	3126	21	A70143	Plasmodium falcipa
c	42	20.8	49.5	3126	21	A70143	Plasmodium falcipa
	43	20.6	49.0	481	21	C94351	Cat flea head and
c	44	20.6	49.0	481	21	C94351	Cat flea head and
	45	20.4	48.6	106	18	T59849	Fragment 105 of PD
c	46	20.4	48.6	106	18	T59849	Fragment 105 of PD
	47	20.4	48.6	237	16	Q81840	PD1 promoter 238 f
c	48	20.4	48.6	237	16	Q81840	PD1 promoter 238 f
	49	20.4	48.6	639	16	Q81838	PD1 promoter 151 f
c	50	20.4	48.6	639	16	Q81838	PD1 promoter 151 f

ALIGNMENTS

```
RESULT   1
A51069
ID   A51069 standard; cDNA to mRNA; 1523 BP.
XX
AC   A51069;
XX
DT   09-OCT-2000  (first entry)
XX
DE   Human LPAAT-gamma-2 coding sequence.
XX
KW   LPAAT-gamma-2; lysophosphatidic acid acyltransferase; cytokine signaling;
KW   1-acyl sn-glycerol-3-phosphate acyltransferase; cytoreductive therapy;
KW   phosphatidic acid metabolism; trilineage haematopoiesis; EC 2.3.1.51;
KW   anti-inflammatory; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..232
FT                   /*tag= a
FT                   /note= "contains two upstream ATGs with in-phase
FT                         stop codons"
FT   CDS             233..1177
FT                   /*tag= b
FT                   /product= LPAAT-gamma-2
FT                   /EC_number= "2.3.1.51"
FT   3'UTR           1178..1523
FT                   /*tag= c
XX
```

```
PN   WO200037655-A1.
XX
PD   29-JUN-2000.
XX
PF   18-DEC-1998;   98WO-US26923.
XX
PR   18-DEC-1998;   98WO-US26923.
XX
PA   (CELL-) CELL THERAPEUTICS INC.
XX
PI   Leung DW,  Adourel D,  Hollenback D;
XX
DR   WPI; 2000-442679/38.
DR   P-PSDB; Y96591.
XX
PT   Novel polynucleotides encoding polypeptides having lysophosphatidic
PT   acid acyltransferase activity for screening compounds for hematopoietic
PT   and anti-inflammatory therapeutic applications
XX
PS   Claim 1; Fig 10; 98pp; English.
XX
CC   Novel isoforms of human lysophosphatidic acid acyltransferase (LPAAT),
CC   also known as 1-acyl sn-glycerol-3-phosphate acyltransferase were
CC   identified by using yeast or plant LPAAT protein sequences for design
CC   of probes. LPAAT is involved in phosphatidic acid metabolism and
CC   signaling in mammalian cells. It was found that overexpression of
CC   LPAAT enhanced the cytokine signaling response in cells. LPAAT is useful
CC   for screening compounds for trilineage haematopoiesis after cytoreductive
CC   therapy or to inhibit inflammation following hypoxia and deoxygenation
CC   injury (e.g. sepsis, trauma and ARDS), and also for developing antibodies
CC   for therapeutic, diagnostic and research use.
XX
SQ   Sequence 1523 BP; 373 A; 409 C; 417 G; 324 T; 0 other;


  Query Match             57.1%;  Score 24;  DB 21;  Length 1523;
  Best Local Similarity   75.0%;  Pred. No. 15;
  Matches   30;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      3 ttctttgctcaaaatttgatcaaattttgagcaaagaata 42
          |||| | || ||| || ||||.|||||| |  ||| || |
Db   1481 ttctctccttaaacttagatcaaattttaaaaaaaaaaaa 1520


RESULT   2
A51069/c
ID   A51069 standard; cDNA to mRNA; 1523 BP.
XX
AC   A51069;
XX
DT   09-OCT-2000  (first entry)
XX
DE   Human LPAAT-gamma-2 coding sequence.
XX
KW   LPAAT-gamma-2; lysophosphatidic acid acyltransferase; cytokine signaling;
KW   1-acyl sn-glycerol-3-phosphate acyltransferase; cytoreductive therapy;
KW   phosphatidic acid metabolism; trilineage haematopoiesis; EC 2.3.1.51;
KW   anti-inflammatory; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..232
FT                   /*tag=  a
FT                   /note= "contains two upstream ATGs with in-phase
FT                         stop codons"
FT   CDS             233..1177
FT                   /*tag=  b
FT                   /product= LPAAT-gamma-2
FT                   /EC_number= "2.3.1.51"
FT   3'UTR           1178..1523
FT                   /*tag=  c
XX
```

```
PN   WO200037655-A1.
XX
PD   29-JUN-2000.
XX
PF   18-DEC-1998;   98WO-US26923.
XX
PR   18-DEC-1998;   98WO-US26923.
XX
PA   (CELL-) CELL THERAPEUTICS INC.
XX
PI   Leung DW,  Adourel D,  Hollenback D;
XX
DR   WPI; 2000-442679/38.
DR   P-PSDB; Y96591.
XX
PT   Novel polynucleotides encoding polypeptides having lysophosphatidic
PT   acid acyltransferase activity for screening compounds for hematopoietic
PT   and anti-inflammatory therapeutic applications
XX
PS   Claim 1; Fig 10; 98pp; English.
XX
CC   Novel isoforms of human lysophosphatidic acid acyltransferase (LPAAT),
CC   also known as 1-acyl sn-glycerol-3-phosphate acyltransferase were
CC   identified by using yeast or plant LPAAT protein sequences for design
CC   of probes. LPAAT is involved in phosphatidic acid metabolism and
CC   signaling in mammalian cells. It was found that overexpression of
CC   LPAAT enhanced the cytokine signaling response in cells. LPAAT is useful
CC   for screening compounds for trilineage haematopoiesis after cytoreductive
CC   therapy or to inhibit inflammation following hypoxia and deoxygenation
CC   injury (e.g. sepsis, trauma and ARDS), and also for developing antibodies
CC   for therapeutic, diagnostic and research use.
XX
SQ   Sequence 1523 BP; 373 A; 409 C; 417 G; 324 T; 0 other;


  Query Match             57.1%;  Score 24;  DB 21;  Length 1523;
  Best Local Similarity   75.0%;  Pred. No. 15;
  Matches   30;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaagaa 40
          | || |||  | ||||||||||| || ||| || | ||||
Db   1520 TTTTTTTTTTTTAAAATTTGATCTAAGTTTAAGGAGAGAA 1481


RESULT   3
A51068
ID   A51068 standard; cDNA to mRNA; 1660 BP.
XX
AC   A51068;
XX
DT   09-OCT-2000  (first entry)
XX
DE   Human LPAAT-gamma-1 coding sequence.
XX
KW   LPAAT-gamma-1; lysophosphatidic acid acyltransferase; cytokine signaling;
KW   1-acyl sn-glycerol-3-phosphate acyltransferase; cytoreductive therapy;
KW   phosphatidic acid metabolism; trilineage haematopoiesis; EC 2.3.1.51;
KW   anti-inflammatory; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..183
FT                   /*tag=  a
FT                   /note= "contains two ATGs and an in-phase stop codon"
FT   CDS             184..1314
FT                   /*tag=  b
FT                   /product= LPAAT-gamma-1
FT                   /EC_number= "2.3.1.51"
FT   3'UTR           1315..1660
FT                   /*tag=  c
XX
PN   WO200037655-A1.
```

```
XX
PD   29-JUN-2000.
XX
PF   18-DEC-1998;   98WO-US26923.
XX
PR   18-DEC-1998;   98WO-US26923.
XX
PA   (CELL-) CELL THERAPEUTICS INC.
XX
PI   Leung DW,  Adourel D,  Hollenback D;
XX
DR   WPI; 2000-442679/38.
DR   P-PSDB; Y96590.
XX
PT   Novel polynucleotides encoding polypeptides having lysophosphatidic
PT   acid acyltransferase activity for screening compounds for hematopoietic
PT   and anti-inflammatory therapeutic applications
XX
PS   Claim 1; Fig 9; 98pp; English.
XX
CC   Novel isoforms of human lysophosphatidic acid acyltransferase (LPAAT),
CC   also known as 1-acyl sn-glycerol-3-phosphate acyltransferase were
CC   identified by using yeast or plant LPAAT protein sequences for design
CC   of probes. LPAAT is involved in phosphatidic acid metabolism and
CC   signaling in mammalian cells. It was found that overexpression of
CC   LPAAT enhanced the cytokine signaling response in cells. LPAAT is useful
CC   for screening compounds for trilineage haematopoiesis after cytoreductive
CC   therapy or to inhibit inflammation following hypoxia and deoxygenation
CC   injury (e.g. sepsis, trauma and ARDS), and also for developing antibodies
CC   for therapeutic, diagnostic and research use.
XX
SQ   Sequence 1660 BP; 367 A; 452 C; 475 G; 366 T; 0 other;


  Query Match             57.1%;  Score 24;  DB 21;  Length 1660;
  Best Local Similarity   75.0%;  Pred. No. 15;
  Matches   30;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       3 ttctttgctcaaaatttgatcaaattttgagcaaagaata 42
           |||| | || ||| || ||||||||||| |  ||| || |
Db    1618 ttctctccttaaacttagatcaaattttaaaaaaaaaaa 1657


RESULT   4
A51068/c
ID   A51068 standard; cDNA to mRNA; 1660 BP.
XX
AC   A51068;
XX
DT   09-OCT-2000  (first entry)
XX
DE   Human LPAAT-gamma-1 coding sequence.
XX
KW   LPAAT-gamma-1; lysophosphatidic acid acyltransferase; cytokine signaling;
KW   1-acyl sn-glycerol-3-phosphate acyltransferase; cytoreductive therapy;
KW   phosphatidic acid metabolism; trilineage haematopoiesis; EC 2.3.1.51;
KW   anti-inflammatory; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..183
FT                   /*tag=  a
FT                   /note= "contains two ATGs and an in-phase stop codon"
FT   CDS             184..1314
FT                   /*tag=  b
FT                   /product= LPAAT-gamma-1
FT                   /EC_number= "2.3.1.51"
FT   3'UTR           1315..1660
FT                   /*tag=  c
XX
PN   WO200037655-A1.
XX
PD   29-JUN-2000.
XX
PF   18-DEC-1998;   98WO-US26923.
XX
PR   18-DEC-1998;   98WO-US26923.
XX
PA   (CELL-) CELL THERAPEUTICS INC.
XX
PI   Leung DW,  Adourel D,  Hollenback D;
XX
DR   WPI; 2000-442679/38.
DR   P-PSDB; Y96590.
XX
PT   Novel polynucleotides encoding polypeptides having lysophosphatidic
PT   acid acyltransferase activity for screening compounds for hematopoietic
PT   and anti-inflammatory therapeutic applications
XX
PS   Claim 1; Fig 9; 98pp; English.
XX
CC   Novel isoforms of human lysophosphatidic acid acyltransferase (LPAAT),
CC   also known as 1-acyl sn-glycerol-3-phosphate acyltransferase were
CC   identified by using yeast or plant LPAAT protein sequences for design
CC   of probes. LPAAT is involved in phosphatidic acid metabolism and
CC   signaling in mammalian cells. It was found that overexpression of
CC   LPAAT enhanced the cytokine signaling response in cells. LPAAT is useful
CC   for screening compounds for trilineage haematopoiesis after cytoreductive
CC   therapy or to inhibit inflammation following hypoxia and deoxygenation
CC   injury (e.g. sepsis, trauma and ARDS), and also for developing antibodies
CC   for therapeutic, diagnostic and research use.
XX
SQ   Sequence 1660 BP; 367 A; 452 C; 475 G; 366 T; 0 other;


  Query Match             57.1%;  Score 24;  DB 21;  Length 1660;
  Best Local Similarity   75.0%;  Pred. No. 15;
  Matches   30;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1 tattctttgctcaaaatttgatcaaattttgagcaaagaa 40
             | || |||  | ||||||||||| || ||| || | ||||
Db    1657 TTTTTTTTTTTTAAAATTTGATCTAAGTTTAAGGAGAGAA 1618


RESULT   5
Z33573
ID   Z33573 standard; cDNA; 2720 BP.
XX
AC   Z33573;
XX
DT   08-DEC-1999  (first entry)
XX
DE   Human breast tumour-associated EST 33.
XX
KW   Expressed sequence tag; EST; human; breast; cancer; cytostatic;
KW   medicaments; gene therapy; treatment; fat metabolism; ss.
XX
OS   Homo sapiens.
XX
PN   DE19813835-A1.
XX
PD   23-SEP-1999.
XX
PF   20-MAR-1998;   98DE-1013835.
XX
PR   20-MAR-1998;   98DE-1013835.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E,  Rosenthal A;
XX
DR   WPI; 1999-528979/45.
XX
PT   Human nucleic acid sequences and protein products from normal breast
PT   tissue, useful for breast cancer therapy  -
```

```
XX
PS   Claim 1a; 121; 206pp; German.
XX
CC   This invention describes novel human nucleic acid sequences from normal
CC   breast tissue which have cytostatic activity. The nucleic acid sequences
CC   can be used to produce and isolate full-length gene sequences. They can
CC   be used to express proteins, which can be used as tools to find an
CC   activity against breast cancer. The sequences can be used in sense or
CC   antisense form. They are especially useful for medicaments for gene
CC   therapy to treat breast cancer and for treating illnesses associated
CC   with fat metabolism. Z33541-Z33610 represent expressed sequence tags
CC   described in the method of the invention.
XX
SQ   Sequence 2720 BP; 947 A; 398 C; 454 G; 921 T; 0 other;

  Query Match             52.4%;  Score 22;  DB 20;  Length 2720;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      19 tgatcaaattttgagcaaagaa 40
           ||||||||||||||||||||||
Db     371 tgatcaaattttgagcaaagaa 392


RESULT   6
Z33573/c
ID   Z33573 standard; cDNA; 2720 BP.
XX
AC   Z33573;
XX
DT   08-DEC-1999  (first entry)
XX
DE   Human breast tumour-associated EST 33.
XX
KW   Expressed sequence tag; EST; human; breast; cancer; cytostatic;
KW   medicaments; gene therapy; treatment; fat metabolism; ss.
XX
OS   Homo sapiens.
XX
PN   DE19813835-A1.
XX
PD   23-SEP-1999.
XX
PF   20-MAR-1998;   98DE-1013835.
XX
PR   20-MAR-1998;   98DE-1013835.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E,  Rosenthal A;
XX
DR   WPI; 1999-528979/45.
XX
PT   Human nucleic acid sequences and protein products from normal breast
PT   tissue, useful for breast cancer therapy  -
XX
PS   Claim 1a; 121; 206pp; German.
XX
CC   This invention describes novel human nucleic acid sequences from normal
CC   breast tissue which have cytostatic activity. The nucleic acid sequences
CC   can be used to produce and isolate full-length gene sequences. They can
CC   be used to express proteins, which can be used as tools to find an
CC   activity against breast cancer. The sequences can be used in sense or
CC   antisense form. They are especially useful for medicaments for gene
CC   therapy to treat breast cancer and for treating illnesses associated
CC   with fat metabolism. Z33541-Z33610 represent expressed sequence tags
CC   described in the method of the invention.
XX
SQ   Sequence 2720 BP; 947 A; 398 C; 454 G; 921 T; 0 other;

  Query Match             52.4%;  Score 22;  DB 20;  Length 2720;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 ttctttgctcaaaatttgatca 24
           ||||||||||||||||||||||
Db     392 TTCTTTGCTCAAAATTTGATCA 371


RESULT   7
X03792
ID   X03792 standard; cDNA; 2975 BP.
XX
AC   X03792;
XX
DT   31-MAR-1999  (first entry)
XX
DE   Human semaphorin E encoding cDNA clone BR5334.
XX
KW   Human; semaphorin E; clone BR5334; nutritional; immune stimulating;
KW   vaccine; haematopoiesis regulating activity; tissue growth;
KW   haemostatic; thrombolytic; anti-inflammatory; tumour inhibition;
KW   gene therapy; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             245..2500
FT                   /*tag=  a
XX
PN   WO9853065-A1.
XX
PD   26-NOV-1998.
XX
PF   19-MAY-1998;   98WO-US10188.
XX
PR   18-MAY-1998;   98US-0080695.
PR   19-MAY-1997;   97US-0858834.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Fechtel K,  Howes SH,  Jacobs K,  Lavallie ER;
PI   McCoy JM,  Merberg D,  Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-059742/05.
DR   P-PSDB; W30617.
XX
PT   New polypeptides encoding secreted human proteins - derived from a
PT   human foetal kidney cDNA library
XX
PS   Claim 1; Page 43-44; 58pp; English.
XX
CC   The present sequence encodes human semaphorin E from cDNA clone BR5334.
CC   Human semaphorin E polynucleotide sequences and protein sequences from
CC   the present invention, are predicted to have biological activities which
CC   would make them suitable for treating, preventing or ameliorating
CC   medical conditions in humans and animals, although no supporting data is
CC   given. Suggested activities include nutritional activity, immune
CC   stimulating (e.g. as vaccines) or suppressing activity, haematopoiesis
CC   regulating activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, and tumour inhibition
CC   activity. The polynucleotide sequences are also stated to be useful for
CC   gene therapy. A host cell transfected with the polynucleotide sequence
CC   encoding human semaphorin E or its subfragments and variants is useful
CC   for recombinant production of the clone BR5334 related protein.
XX
SQ   Sequence 2975 BP; 887 A; 620 C; 665 G; 802 T; 1 other;

  Query Match             52.4%;  Score 22;  DB 20;  Length 2975;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      19 tgatcaaattttgagcaaagaa 40
           ||||||||||||||||||||||
Db    2552 tgatcaaattttgagcaaagaa 2573


RESULT   8
X03792/c
ID   X03792 standard; cDNA; 2975 BP.
XX
AC   X03792;
XX
DT   31-MAR-1999  (first entry)
XX
DE   Human semaphorin E encoding cDNA clone BR5334.
XX
KW   Human; semaphorin E; clone BR5334; nutritional; immune stimulating;
KW   vaccine; haematopoiesis regulating activity; tissue growth;
KW   haemostatic; thrombolytic; anti-inflammatory; tumour inhibition;
KW   gene therapy; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             245..2500
FT                   /*tag=  a
XX
PN   WO9853065-A1.
XX
PD   26-NOV-1998.
XX
PF   19-MAY-1998;    98WO-US10188.
XX
PR   18-MAY-1998;    98US-0080695.
PR   19-MAY-1997;    97US-0858834.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Fechtel K,  Howes SH,  Jacobs K,  Lavallie ER;
PI   McCoy JM,  Merberg D,  Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-059742/05.
DR   P-PSDB; W30617.
XX
PT   New polypeptides encoding secreted human proteins - derived from a
PT   human foetal kidney cDNA library
XX
PS   Claim 1; Page 43-44; 58pp; English.
XX
CC   The present sequence encodes human semaphorin E from cDNA clone BR5334.
CC   Human semaphorin E polynucleotide sequences and protein sequences from
CC   the present invention, are predicted to have biological activities which
CC   would make them suitable for treating, preventing or ameliorating
CC   medical conditions in humans and animals, although no supporting data is
CC   given. Suggested activities include nutritional activity, immune
CC   stimulating (e.g. as vaccines) or suppressing activity, haematopoiesis
CC   regulating activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, and tumour inhibition
CC   activity. The polynucleotide sequences are also stated to be useful for
CC   gene therapy. A host cell transfected with the polynucleotide sequence
CC   encoding human semaphorin E or its subfragments and variants is useful
CC   for recombinant production of the clone BR5334 related protein.
XX
SQ   Sequence 2975 BP; 887 A; 620 C; 665 G; 802 T; 1 other;


  Query Match             52.4%;  Score 22;  DB 20;  Length 2975;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 ttctttgctcaaaatttgatca 24
           ||||||||||||||||||||||
Db    2573 TTCTTTGCTCAAAATTTGATCA 2552


RESULT   9
C66800
ID   C66800 standard; cDNA; 2975 BP.
XX
AC   C66800;
XX
DT   19-FEB-2001  (first entry)
XX
DE   Clone BR533_4 coding sequence.
XX
KW   Human; clone BR533_4; secreted protein; immune deficiency;
KW   microbial infection; autoimmune disorder; allergy; asthma; inflammation;
KW   myeloid deficiency; lymphoid cell deficiency; anaemia; burn;
KW   wound healing; ulcer; periodontal disease; nervous system disease;
KW   neuropathy; lung fibrosis; liver fibrosis; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200063692-A1.
XX
PD   26-OCT-2000.
XX
PF   14-APR-2000; 2000WO-US10048.
XX
PR   15-APR-1999;   99US-0292550.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  McCoy JM,  Lavallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M,  Spaulding V;
XX
DR   WPI; 2000-679620/66.
DR   P-PSDB; B28379.
XX
PT   New monoclonal antibodies, useful for treating cancer and
PT   immunodetection of secreted proteins which are in turn useful for
PT   treating neurological, inflammatory, immune diseases and microbial
PT   infections  -
XX
PS   Disclosure; Pages 66-67; 75pp; English.
XX
CC   The present sequence is the coding sequence for human clone BR533_4. The
CC   present sequence was isolated from a human foetal kidney cDNA library.
CC   The BR533_4 protein is a secreted protein and can be used to treat a
CC   number of conditions including various immune deficiencies and disorders,
CC   microbial infections, autoimmune disorders, allergic reactions such as
CC   asthma, respiratory problems, inflammation, myeloid or lymphoid cell
CC   deficiencies, anaemias, burns, wound healing, ulcers, periodontal
CC   disease, central and peripheral nervous system diseases and neuropathies,
CC   lung or liver fibrosis and cancer.
XX
SQ   Sequence 2975 BP; 887 A; 620 C; 665 G; 802 T; 1 other;


  Query Match             52.4%;  Score 22;  DB 21;  Length 2975;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      19 tgatcaaattttgagcaaagaa 40
           ||||||||||||||||||||||
Db    2552 tgatcaaattttgagcaaagaa 2573


RESULT  10
C66800/c
ID   C66800 standard; cDNA; 2975 BP.
XX
AC   C66800;
XX
```

DT 19-FEB-2001 (first entry)
XX
DE Clone BR533_4 coding sequence.
XX
KW Human; clone BR533_4; secreted protein; immune deficiency;
KW microbial infection; autoimmune disorder; allergy; asthma; inflammation;
KW myeloid deficiency; lymphoid cell deficiency; anaemia; burn;
KW wound healing; ulcer; periodontal disease; nervous system disease;
KW neuropathy; lung fibrosis; liver fibrosis; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200063692-A1.
XX
PD 26-OCT-2000.
XX
PF 14-APR-2000; 2000WO-US10048.
XX
PR 15-APR-1999; 99US-0292550.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Spaulding V;
XX
DR WPI; 2000-679620/66.
DR P-PSDB; B28379.
XX
PT New monoclonal antibodies, useful for treating cancer and
PT immunodetection of secreted proteins which are in turn useful for
PT treating neurological, inflammatory, immune diseases and microbial
PT infections -
XX
PS Disclosure; Pages 66-67; 75pp; English.
XX
CC The present sequence is the coding sequence for human clone BR533_4. The
CC present sequence was isolated from a human foetal kidney cDNA library.
CC The BR533_4 protein is a secreted protein and can be used to treat a
CC number of conditions including various immune deficiencies and disorders,
CC microbial infections, autoimmune disorders, allergic reactions such as
CC asthma, respiratory problems, inflammation, myeloid or lymphoid cell
CC deficiencies, anaemias, burns, wound healing, ulcers, periodontal
CC disease, central and peripheral nervous system diseases and neuropathies,
CC lung or liver fibrosis and cancer.
XX
SQ Sequence 2975 BP; 887 A; 620 C; 665 G; 802 T; 1 other;

Query Match 52.4%; Score 22; DB 21; Length 2975;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ttctttgctcaaaatttgatca 24
||||||||||||||||||||||
Db 2573 TTCTTTGCTCAAAATTTGATCA 2552

RESULT 11
V68977
ID V68977 standard; DNA; 715 BP.
XX
AC V68977;
XX
DT 22-JAN-1999 (first entry)
XX
DE DNA molecule encoding a breast tumour specific polypeptide #169.
XX
KW Human; breast cancer; breast tumour tissue; diagnosis; treatment;
KW vaccine; epitope; endogenous; retroviral element; ss.
XX
OS Homo sapiens.
XX
PN WO9845328-A2.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-US06939.
XX
PR 11-DEC-1997; 97US-0991789.
PR 09-APR-1997; 97US-0838762.
XX
PA (CORI-) CORIXA CORP.
XX
PI Frudakis TN, Reed SG, Smith JM;
XX
DR WPI; 1998-557473/47.
XX
PT New DNA sequences isolated from endogenous human retroviral element
PT - and related vectors, transformed cells, proteins and antibodies,
PT useful for diagnosis, treatment and prevention of breast cancer
XX
PS Claim 1; Page 127; 173pp; English.
XX
CC V68800 to V68998 represent nucleotide sequences which encode human
CC breast tumour specific polypeptides. Detection or measurement of
CC human breast tumour specific polypeptides and nucleotide sequences,
CC or the corresponding RNA in a sample, is used for diagnosis and
CC monitoring of breast cancer. Human breast tumour specific polypeptides
CC and nucleotide sequences, and the vectors containing the DNAs, are also
CC useful in vaccines for inhibiting development (for prevention or
CC therapy) of breast cancer. The polypeptides may also be used to
CC raise monoclonal antibodies, used as immunoassay reagents.
XX
SQ Sequence 715 BP; 183 A; 131 C; 86 G; 294 T; 21 other;

Query Match 51.4%; Score 21.6; DB 19; Length 715;
Best Local Similarity 73.0%; Pred. No. 88;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 4 tctttgctcaaaatttgatcaaattttgagcaaagaa 40
| |||||| ||||| ||||||| || ||| |||
Db 418 tttttgcttaaaatgtgatcaatattccttcaangaa 454

RESULT 12
V68977/c
ID V68977 standard; DNA; 715 BP.
XX
AC V68977;
XX
DT 22-JAN-1999 (first entry)
XX
DE DNA molecule encoding a breast tumour specific polypeptide #169.
XX
KW Human; breast cancer; breast tumour tissue; diagnosis; treatment;
KW vaccine; epitope; endogenous; retroviral element; ss.
XX
OS Homo sapiens.
XX
PN WO9845328-A2.
XX
PD 15-OCT-1998.
XX
PF 09-APR-1998; 98WO-US06939.
XX
PR 11-DEC-1997; 97US-0991789.
PR 09-APR-1997; 97US-0838762.
XX
PA (CORI-) CORIXA CORP.
XX
PI Frudakis TN, Reed SG, Smith JM;
XX
DR WPI; 1998-557473/47.
XX
PT New DNA sequences isolated from endogenous human retroviral element

```
PT   - and related vectors, transformed cells, proteins and antibodies,
PT   useful for diagnosis, treatment and prevention of breast cancer
XX
PS   Claim 1; Page 127; 173pp; English.
XX
CC   V68800 to V68998 represent nucleotide sequences which encode human
CC   breast tumour specific polypeptides. Detection or measurement of
CC   human breast tumour specific polypeptides and nucleotide sequences,
CC   or the corresponding RNA in a sample, is used for diagnosis and
CC   monitoring of breast cancer. Human breast tumour specific polypeptides
CC   and nucleotide sequences, and the vectors containing the DNAs, are also
CC   useful in vaccines for inhibiting development (for prevention or
CC   therapy) of breast cancer. The polypeptides may also be used to
CC   raise monoclonal antibodies, used as immunoassay reagents.
XX
SQ   Sequence 715 BP; 183 A; 131 C; 86 G; 294 T; 21 other;


  Query Match             51.4%;  Score 21.6;  DB 19;  Length 715;
  Best Local Similarity   73.0%;  Pred. No. 88;
  Matches   27;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       3 ttctttgctcaaaatttgatcaaattttgagcaaaga 39
           ||| |||    ||  ||||||| ||||| ||||||
Db     454 TTCNTTGAAGGAATATTGATCACATTTTAAGCAAAAA 418


RESULT  13
C80976
ID   C80976 standard; cDNA; 715 BP.
XX
AC   C80976;
XX
DT   13-FEB-2001  (first entry)
XX
DE   Human breast tumour-specific cDNA SEQ ID NO: 264.
XX
KW   Human; breast tumour-specific antigen; cytostatic; vaccine;
KW   breast cancer; B18Ag1; B11Ag1; B15Ag1; ss.
XX
OS   Homo sapiens.
XX
PN   WO200061753-A2.
XX
PD   19-OCT-2000.
XX
PF   07-APR-2000; 2000WO-US09312.
XX
PR   09-APR-1999;   99US-0289198.
PR   28-OCT-1999;   99US-0429755.
PR   23-MAR-2000; 2000US-0534825.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Frudakis TN,  Smith JM,  Reed SG,  Misher LE,  Retter MW,  Dillon DC;
XX
DR   WPI; 2000-628403/60.
XX
PT   An isolated polypeptide comprising an immunogenic portion of a breast
PT   tumor protein used for inhibiting the development of cancer, especially
PT   breast cancer, and monitoring cancer progression in a patient -
XX
PS   Claim 4; Page 162-163; 187pp; English.
XX
CC   The present sequence is given in a specification relating to compositions
CC   and methods for the treatment and diagnosis of breast cancer. Nucleotide
CC   sequences that are preferentially expressed in breast tumour tissue, and
CC   the polypeptides encoded by such nucleotide sequences, are used in
CC   compositions and vaccines to inhibit the development of cancer,
CC   especially breast cancer. The progression of a cancer may be monitored by
CC   carrying out detection of tumour-specific antigens at subsequent time
CC   points and comparing the results from the different time points.
CC   CD4+ and/or CD8+ T-Cells isolated from the cancer patient may be treated
CC   with tumour-specific polypeptides, polynucleotides encoding the
CC   polypeptides or antigen presenting cells expressing the polypeptides. The
CC   cells are then administered to the patient to inhibit development of
CC   cancer.
XX
SQ   Sequence 715 BP; 183 A; 131 C; 86 G; 294 T; 21 other;


  Query Match             51.4%;  Score 21.6;  DB 21;  Length 715;
  Best Local Similarity   73.0%;  Pred. No. 88;
  Matches   27;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       4 tctttgctcaaaatttgatcaaattttgagcaaagaa 40
           | |||||| ||||| |||||||  ||    ||| |||
Db     418 tttttgcttaaaatgtgatcaatattccttcaangaa 454


RESULT  14
C80976/c
ID   C80976 standard; cDNA; 715 BP.
XX
AC   C80976;
XX
DT   13-FEB-2001  (first entry)
XX
DE   Human breast tumour-specific cDNA SEQ ID NO: 264.
XX
KW   Human; breast tumour-specific antigen; cytostatic; vaccine;
KW   breast cancer; B18Ag1; B11Ag1; B15Ag1; ss.
XX
OS   Homo sapiens.
XX
PN   WO200061753-A2.
XX
PD   19-OCT-2000.
XX
PF   07-APR-2000; 2000WO-US09312.
XX
PR   09-APR-1999;   99US-0289198.
PR   28-OCT-1999;   99US-0429755.
PR   23-MAR-2000; 2000US-0534825.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Frudakis TN,  Smith JM,  Reed SG,  Misher LE,  Retter MW,  Dillon DC;
XX
DR   WPI; 2000-628403/60.
XX
PT   An isolated polypeptide comprising an immunogenic portion of a breast
PT   tumor protein used for inhibiting the development of cancer, especially
PT   breast cancer, and monitoring cancer progression in a patient -
XX
PS   Claim 4; Page 162-163; 187pp; English.
XX
CC   The present sequence is given in a specification relating to compositions
CC   and methods for the treatment and diagnosis of breast cancer. Nucleotide
CC   sequences that are preferentially expressed in breast tumour tissue, and
CC   the polypeptides encoded by such nucleotide sequences, are used in
CC   compositions and vaccines to inhibit the development of cancer,
CC   especially breast cancer. The progression of a cancer may be monitored by
CC   carrying out detection of tumour-specific antigens at subsequent time
CC   points and comparing the results from the different time points.
CC   CD4+ and/or CD8+ T-Cells isolated from the cancer patient may be treated
CC   with tumour-specific polypeptides, polynucleotides encoding the
CC   polypeptides or antigen presenting cells expressing the polypeptides. The
CC   cells are then administered to the patient to inhibit development of
CC   cancer.
XX
SQ   Sequence 715 BP; 183 A; 131 C; 86 G; 294 T; 21 other;


  Query Match             51.4%;  Score 21.6;  DB 21;  Length 715;
  Best Local Similarity   73.0%;  Pred. No. 88;
```

```
  Matches   27;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      3 ttctttgctcaaaatttgatcaaattttgagcaaaga 39
          ||| |||    ||  ||||||| ||||| |||||| |
Db    454 TTCNTTGAAGGAATATTGATCACATTTTAAGCAAAAA 418


RESULT  15
V59767
ID   V59767 standard; DNA; 1086 BP.
XX
AC   V59767;
XX
DT   19-JAN-1999  (first entry)
XX
DE   Human secreted protein gene 108 clone HEBEK93.
XX
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS   Homo sapiens.
XX
PN   WO9839448-A2.
XX
PD   11-SEP-1998.
XX
PF   06-MAR-1998;   98WO-US04493.
XX
PR   02-OCT-1997;   97US-0061060.
PR   07-MAR-1997;   97US-0038621.
PR   07-MAR-1997;   97US-0040161.
PR   07-MAR-1997;   97US-0040162.
PR   07-MAR-1997;   97US-0040163.
PR   07-MAR-1997;   97US-0040333.
PR   07-MAR-1997;   97US-0040334.
PR   07-MAR-1997;   97US-0040336.
PR   07-MAR-1997;   97US-0040626.
PR   11-APR-1997;   97US-0043311.
PR   11-APR-1997;   97US-0043312.
PR   11-APR-1997;   97US-0043313.
PR   11-APR-1997;   97US-0043314.
PR   11-APR-1997;   97US-0043568.
PR   11-APR-1997;   97US-0043569.
PR   11-APR-1997;   97US-0043576.
PR   11-APR-1997;   97US-0043578.
PR   11-APR-1997;   97US-0043580.
PR   11-APR-1997;   97US-0043669.
PR   11-APR-1997;   97US-0043670.
PR   11-APR-1997;   97US-0043671.
PR   11-APR-1997;   97US-0043672.
PR   11-APR-1997;   97US-0043674.
PR   23-MAY-1997;   97US-0047492.
PR   23-MAY-1997;   97US-0047500.
PR   23-MAY-1997;   97US-0047501.
PR   23-MAY-1997;   97US-0047502.
PR   23-MAY-1997;   97US-0047503.
PR   23-MAY-1997;   97US-0047581.
PR   23-MAY-1997;   97US-0047582.
PR   23-MAY-1997;   97US-0047583.
PR   23-MAY-1997;   97US-0047584.
PR   23-MAY-1997;   97US-0047585.
PR   23-MAY-1997;   97US-0047586.
PR   23-MAY-1997;   97US-0047587.
PR   23-MAY-1997;   97US-0047588.
PR   23-MAY-1997;   97US-0047589.
PR   23-MAY-1997;   97US-0047590.
PR   23-MAY-1997;   97US-0047592.
PR   23-MAY-1997;   97US-0047593.
PR   23-MAY-1997;   97US-0047594.
PR   23-MAY-1997;   97US-0047595.
PR   23-MAY-1997;   97US-0047596.
PR   23-MAY-1997;   97US-0047597.
PR   23-MAY-1997;   97US-0047598.
PR   23-MAY-1997;   97US-0047599.
PR   23-MAY-1997;   97US-0047600.
PR   23-MAY-1997;   97US-0047601.
PR   23-MAY-1997;   97US-0047612.
PR   23-MAY-1997;   97US-0047613.
PR   23-MAY-1997;   97US-0047614.
PR   23-MAY-1997;   97US-0047615.
PR   23-MAY-1997;   97US-0047617.
PR   23-MAY-1997;   97US-0047618.
PR   23-MAY-1997;   97US-0047632.
PR   23-MAY-1997;   97US-0047633.
PR   06-JUN-1997;   97US-0048964.
PR   06-JUN-1997;   97US-0048974.
PR   13-JUN-1997;   97US-0049610.
PR   08-JUL-1997;   97US-0051926.
PR   16-JUL-1997;   97US-0052874.
PR   18-AUG-1997;   97US-0055724.
PR   22-AUG-1997;   97US-0056630.
PR   22-AUG-1997;   97US-0056631.
PR   22-AUG-1997;   97US-0056632.
PR   22-AUG-1997;   97US-0056636.
PR   22-AUG-1997;   97US-0056637.
PR   22-AUG-1997;   97US-0056662.
PR   22-AUG-1997;   97US-0056664.
PR   22-AUG-1997;   97US-0056845.
PR   22-AUG-1997;   97US-0056862.
PR   22-AUG-1997;   97US-0056864.
PR   22-AUG-1997;   97US-0056872.
PR   22-AUG-1997;   97US-0056874.
PR   22-AUG-1997;   97US-0056875.
PR   22-AUG-1997;   97US-0056876.
PR   22-AUG-1997;   97US-0056877.
PR   22-AUG-1997;   97US-0056878.
PR   22-AUG-1997;   97US-0056879.
PR   22-AUG-1997;   97US-0056880.
PR   22-AUG-1997;   97US-0056881.
PR   22-AUG-1997;   97US-0056882.
PR   22-AUG-1997;   97US-0056884.
PR   22-AUG-1997;   97US-0056886.
PR   22-AUG-1997;   97US-0056887.
PR   22-AUG-1997;   97US-0056888.
PR   22-AUG-1997;   97US-0056889.
PR   22-AUG-1997;   97US-0056892.
PR   22-AUG-1997;   97US-0056893.
PR   22-AUG-1997;   97US-0056894.
PR   22-AUG-1997;   97US-0056903.
PR   22-AUG-1997;   97US-0056908.
PR   22-AUG-1997;   97US-0056909.
PR   22-AUG-1997;   97US-0056910.
PR   22-AUG-1997;   97US-0056911.
PR   05-SEP-1997;   97US-0057650.
PR   05-SEP-1997;   97US-0057669.
PR   05-SEP-1997;   97US-0057761.
PR   12-SEP-1997;   97US-0058785.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Bednarik DP,  Brewer LA,  Carter KC,  Duan R,  Ebner R,  Endress GA;
PI   Feng P,  Ferrie AM,  Fischer CL,  Florence KA,  Greene JM,  Hu JS;
PI   Kyaw H,  Lafleur DW,  Li Y,  Moore PA,  Ni J,  Olsen HS,  Rosen CA;
PI   Ruben SM,  Shi Y,  Soppet DR,  Young PE,  Yu GL,  Zeng Z;
XX
DR   WPI; 1998-506364/43.
DR   P-PSDB; W74982.
XX
PT   New isolated human genes and the secreted polypeptide(s) they encode
PT   - useful for diagnosis and treatment of e.g. cancers, neurological
```

```
PT   disorders, immune diseases, inflammation or blood disorders
XX
PS   Claim 1; Page 494; 721pp; English.
XX
CC   This sequence represents a nucleic acid molecule designated Gene 108 from
CC   the human cDNA clone HEBEK93 (deposited as clone ATCC 97901 and ATCC
CC   209047) which encodes a secreted human protein.  The gene can be used to
CC   generate fusion proteins by linking to the gene to a human immunoglobulin
CC   Fc portion (e.g. V59502) for increasing the stability of the fused
CC   protein as compared to the human protein only.
CC   The invention relates to 186 novel genes and their fragments (nucleic
CC   acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which
CC   are useful for preventing, treating or ameliorating  medical conditions
CC   e.g. by protein or gene therapy.  Also, pathological conditions can be
CC   diagnosed by determining the amount of the new polypeptides in a sample
CC   or by determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 186 polynucleotides, based on
CC   which tissues they are most highly expressed in (see V59511 for described
CC   uses).
XX
SQ   Sequence 1086 BP; 315 A; 243 C; 266 G; 259 T; 3 other;


  Query Match             51.0%;  Score 21.4;  DB 19;  Length 1086;
  Best Local Similarity   71.8%;  Pred. No. 1.1e+02;
  Matches   28;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       4 tctttgctcaaaatttgatcaaattttgagcaaagaata 42
           ||  |||||  || ||| |||||| |  ||||||| | |
Db     571 tcggtgctcggaacttgctcaaatctatagcaaagcaga 609


RESULT  16
V59767/c
ID   V59767 standard; DNA; 1086 BP.
XX
AC   V59767;
XX
DT   19-JAN-1999  (first entry)
XX
DE   Human secreted protein gene 108 clone HEBEK93.
XX
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS   Homo sapiens.
XX
PN   WO9839448-A2.
XX
PD   11-SEP-1998.
XX
PF   06-MAR-1998;   98WO-US04493.
XX
PR   02-OCT-1997;   97US-0061060.
PR   07-MAR-1997;   97US-0038621.
PR   07-MAR-1997;   97US-0040161.
PR   07-MAR-1997;   97US-0040162.
PR   07-MAR-1997;   97US-0040163.
PR   07-MAR-1997;   97US-0040333.
PR   07-MAR-1997;   97US-0040334.
PR   07-MAR-1997;   97US-0040336.
PR   07-MAR-1997;   97US-0040626.
PR   11-APR-1997;   97US-0043311.
PR   11-APR-1997;   97US-0043312.
PR   11-APR-1997;   97US-0043313.
PR   11-APR-1997;   97US-0043314.
PR   11-APR-1997;   97US-0043568.
PR   11-APR-1997;   97US-0043569.
PR   11-APR-1997;   97US-0043576.
PR   11-APR-1997;   97US-0043578.
PR   11-APR-1997;   97US-0043580.
PR   11-APR-1997;   97US-0043669.
PR   11-APR-1997;   97US-0043670.
PR   11-APR-1997;   97US-0043671.
PR   11-APR-1997;   97US-0043672.
PR   11-APR-1997;   97US-0043674.
PR   23-MAY-1997;   97US-0047492.
PR   23-MAY-1997;   97US-0047500.
PR   23-MAY-1997;   97US-0047501.
PR   23-MAY-1997;   97US-0047502.
PR   23-MAY-1997;   97US-0047503.
PR   23-MAY-1997;   97US-0047581.
PR   23-MAY-1997;   97US-0047582.
PR   23-MAY-1997;   97US-0047583.
PR   23-MAY-1997;   97US-0047584.
PR   23-MAY-1997;   97US-0047585.
PR   23-MAY-1997;   97US-0047586.
PR   23-MAY-1997;   97US-0047587.
PR   23-MAY-1997;   97US-0047588.
PR   23-MAY-1997;   97US-0047589.
PR   23-MAY-1997;   97US-0047590.
PR   23-MAY-1997;   97US-0047592.
PR   23-MAY-1997;   97US-0047593.
PR   23-MAY-1997;   97US-0047594.
PR   23-MAY-1997;   97US-0047595.
PR   23-MAY-1997;   97US-0047596.
PR   23-MAY-1997;   97US-0047597.
PR   23-MAY-1997;   97US-0047598.
PR   23-MAY-1997;   97US-0047599.
PR   23-MAY-1997;   97US-0047600.
PR   23-MAY-1997;   97US-0047601.
PR   23-MAY-1997;   97US-0047612.
PR   23-MAY-1997;   97US-0047613.
PR   23-MAY-1997;   97US-0047614.
PR   23-MAY-1997;   97US-0047615.
PR   23-MAY-1997;   97US-0047617.
PR   23-MAY-1997;   97US-0047618.
PR   23-MAY-1997;   97US-0047632.
PR   23-MAY-1997;   97US-0047633.
PR   06-JUN-1997;   97US-0048964.
PR   06-JUN-1997;   97US-0048974.
PR   13-JUN-1997;   97US-0049610.
PR   08-JUL-1997;   97US-0051926.
PR   16-JUL-1997;   97US-0052874.
PR   18-AUG-1997;   97US-0055724.
PR   22-AUG-1997;   97US-0056630.
PR   22-AUG-1997;   97US-0056631.
PR   22-AUG-1997;   97US-0056632.
PR   22-AUG-1997;   97US-0056636.
PR   22-AUG-1997;   97US-0056637.
PR   22-AUG-1997;   97US-0056662.
PR   22-AUG-1997;   97US-0056664.
PR   22-AUG-1997;   97US-0056845.
PR   22-AUG-1997;   97US-0056862.
PR   22-AUG-1997;   97US-0056864.
PR   22-AUG-1997;   97US-0056872.
PR   22-AUG-1997;   97US-0056874.
PR   22-AUG-1997;   97US-0056875.
PR   22-AUG-1997;   97US-0056876.
PR   22-AUG-1997;   97US-0056877.
PR   22-AUG-1997;   97US-0056878.
PR   22-AUG-1997;   97US-0056879.
PR   22-AUG-1997;   97US-0056880.
PR   22-AUG-1997;   97US-0056881.
PR   22-AUG-1997;   97US-0056882.
PR   22-AUG-1997;   97US-0056884.
PR   22-AUG-1997;   97US-0056886.
PR   22-AUG-1997;   97US-0056887.
PR   22-AUG-1997;   97US-0056888.
PR   22-AUG-1997;   97US-0056889.
```

```
PR   22-AUG-1997;   97US-0056892.
PR   22-AUG-1997;   97US-0056893.
PR   22-AUG-1997;   97US-0056894.
PR   22-AUG-1997;   97US-0056903.
PR   22-AUG-1997;   97US-0056908.
PR   22-AUG-1997;   97US-0056909.
PR   22-AUG-1997;   97US-0056910.
PR   22-AUG-1997;   97US-0056911.
PR   05-SEP-1997;   97US-0057650.
PR   05-SEP-1997;   97US-0057669.
PR   05-SEP-1997;   97US-0057761.
PR   12-SEP-1997;   97US-0058785.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;
PI   Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
PI   Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
PI   Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
XX
DR   WPI; 1998-506364/43.
DR   P-PSDB; W74982.
XX
PT   New isolated human genes and the secreted polypeptide(s) they encode
PT   - useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
XX
PS   Claim 1; Page 494; 721pp; English.
XX
CC   This sequence represents a nucleic acid molecule designated Gene 108 from
CC   the human cDNA clone HEBEK93 (deposited as clone ATCC 97901 and ATCC
CC   209047) which encodes a secreted human protein. The gene can be used to
CC   generate fusion proteins by linking to the gene to a human immunoglobulin
CC   Fc portion (e.g. V59502) for increasing the stability of the fused
CC   protein as compared to the human protein only.
CC   The invention relates to 186 novel genes and their fragments (nucleic
CC   acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which
CC   are useful for preventing, treating or ameliorating medical conditions
CC   e.g. by protein or gene therapy. Also, pathological conditions can be
CC   diagnosed by determining the amount of the new polypeptides in a sample
CC   or by determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 186 polynucleotides, based on
CC   which tissues they are most highly expressed in (see V59511 for described
CC   uses).
XX
SQ   Sequence 1086 BP; 315 A; 243 C; 266 G; 259 T; 3 other;


  Query Match             51.0%;  Score 21.4;  DB 19;  Length 1086;
  Best Local Similarity   71.8%;  Pred. No. 1.1e+02;
  Matches   28;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1 tattctttgctcaaaatttgatcaaattttgagcaaaga 39
           | | |||||||  | ||||||  ||| ||  |||||  ||
Db     609 TCTGCTTTGCTATAGATTTGAGCAAGTTCCGAGCACCGA 571


RESULT  17
V59618
ID   V59618 standard; DNA; 1256 BP.
XX
AC   V59618;
XX
DT   19-JAN-1999  (first entry)
XX
DE   Human secreted protein gene 108 clone HEBEK93.
XX
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS   Homo sapiens.
XX
PN   WO9839448-A2.
XX
PD   11-SEP-1998.
XX
PF   06-MAR-1998;   98WO-US04493.
XX
PR   02-OCT-1997;   97US-0061060.
PR   07-MAR-1997;   97US-0038621.
PR   07-MAR-1997;   97US-0040161.
PR   07-MAR-1997;   97US-0040162.
PR   07-MAR-1997;   97US-0040163.
PR   07-MAR-1997;   97US-0040333.
PR   07-MAR-1997;   97US-0040334.
PR   07-MAR-1997;   97US-0040336.
PR   07-MAR-1997;   97US-0040626.
PR   11-APR-1997;   97US-0043311.
PR   11-APR-1997;   97US-0043312.
PR   11-APR-1997;   97US-0043313.
PR   11-APR-1997;   97US-0043314.
PR   11-APR-1997;   97US-0043568.
PR   11-APR-1997;   97US-0043569.
PR   11-APR-1997;   97US-0043576.
PR   11-APR-1997;   97US-0043578.
PR   11-APR-1997;   97US-0043580.
PR   11-APR-1997;   97US-0043669.
PR   11-APR-1997;   97US-0043670.
PR   11-APR-1997;   97US-0043671.
PR   11-APR-1997;   97US-0043672.
PR   11-APR-1997;   97US-0043674.
PR   23-MAY-1997;   97US-0047492.
PR   23-MAY-1997;   97US-0047500.
PR   23-MAY-1997;   97US-0047501.
PR   23-MAY-1997;   97US-0047502.
PR   23-MAY-1997;   97US-0047503.
PR   23-MAY-1997;   97US-0047581.
PR   23-MAY-1997;   97US-0047582.
PR   23-MAY-1997;   97US-0047583.
PR   23-MAY-1997;   97US-0047584.
PR   23-MAY-1997;   97US-0047585.
PR   23-MAY-1997;   97US-0047586.
PR   23-MAY-1997;   97US-0047587.
PR   23-MAY-1997;   97US-0047588.
PR   23-MAY-1997;   97US-0047589.
PR   23-MAY-1997;   97US-0047590.
PR   23-MAY-1997;   97US-0047592.
PR   23-MAY-1997;   97US-0047593.
PR   23-MAY-1997;   97US-0047594.
PR   23-MAY-1997;   97US-0047595.
PR   23-MAY-1997;   97US-0047596.
PR   23-MAY-1997;   97US-0047597.
PR   23-MAY-1997;   97US-0047598.
PR   23-MAY-1997;   97US-0047599.
PR   23-MAY-1997;   97US-0047600.
PR   23-MAY-1997;   97US-0047601.
PR   23-MAY-1997;   97US-0047612.
PR   23-MAY-1997;   97US-0047613.
PR   23-MAY-1997;   97US-0047614.
PR   23-MAY-1997;   97US-0047615.
PR   23-MAY-1997;   97US-0047617.
PR   23-MAY-1997;   97US-0047618.
PR   23-MAY-1997;   97US-0047632.
PR   23-MAY-1997;   97US-0047633.
PR   06-JUN-1997;   97US-0048964.
PR   06-JUN-1997;   97US-0048974.
PR   13-JUN-1997;   97US-0049610.
PR   08-JUL-1997;   97US-0051926.
PR   16-JUL-1997;   97US-0052874.
PR   18-AUG-1997;   97US-0055724.
```

```
PR   22-AUG-1997;   97US-0056630.
PR   22-AUG-1997;   97US-0056631.
PR   22-AUG-1997;   97US-0056632.
PR   22-AUG-1997;   97US-0056636.
PR   22-AUG-1997;   97US-0056637.
PR   22-AUG-1997;   97US-0056662.
PR   22-AUG-1997;   97US-0056664.
PR   22-AUG-1997;   97US-0056845.
PR   22-AUG-1997;   97US-0056862.
PR   22-AUG-1997;   97US-0056864.
PR   22-AUG-1997;   97US-0056872.
PR   22-AUG-1997;   97US-0056874.
PR   22-AUG-1997;   97US-0056875.
PR   22-AUG-1997;   97US-0056876.
PR   22-AUG-1997;   97US-0056877.
PR   22-AUG-1997;   97US-0056878.
PR   22-AUG-1997;   97US-0056879.
PR   22-AUG-1997;   97US-0056880.
PR   22-AUG-1997;   97US-0056881.
PR   22-AUG-1997;   97US-0056882.
PR   22-AUG-1997;   97US-0056884.
PR   22-AUG-1997;   97US-0056886.
PR   22-AUG-1997;   97US-0056887.
PR   22-AUG-1997;   97US-0056888.
PR   22-AUG-1997;   97US-0056889.
PR   22-AUG-1997;   97US-0056892.
PR   22-AUG-1997;   97US-0056893.
PR   22-AUG-1997;   97US-0056894.
PR   22-AUG-1997;   97US-0056903.
PR   22-AUG-1997;   97US-0056908.
PR   22-AUG-1997;   97US-0056909.
PR   22-AUG-1997;   97US-0056910.
PR   22-AUG-1997;   97US-0056911.
PR   05-SEP-1997;   97US-0057650.
PR   05-SEP-1997;   97US-0057669.
PR   05-SEP-1997;   97US-0057761.
PR   12-SEP-1997;   97US-0058785.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Bednarik DP,  Brewer LA,  Carter KC,  Duan R,  Ebner R,  Endress GA;
PI   Feng P,  Ferrie AM,  Fischer CL,  Florence KA,  Greene JM,  Hu JS;
PI   Kyaw H,  Lafleur DW,  Li Y,  Moore PA,  Ni J,  Olsen HS,  Rosen CA;
PI   Ruben SM,  Shi Y,  Soppet DR,  Young PE,  Yu GL,  Zeng Z;
XX
DR   WPI; 1998-506364/43.
DR   P-PSDB; W74836.
XX
PT   New isolated human genes and the secreted polypeptide(s) they encode
PT   - useful for diagnosis and treatment of e.g. cancers, neurological
PT   disorders, immune diseases, inflammation or blood disorders
XX
PS   Claim 1; Page 340-341; 721pp; English.
XX
CC   This sequence represents a nucleic acid molecule designated Gene 108 from
CC   the human cDNA clone HEBEK93 (deposited as clone ATCC 97901 and ATCC
CC   209047) which encodes a secreted human protein.  The gene can be used to
CC   generate fusion proteins by linking to the gene to a human immunoglobulin
CC   Fc portion (e.g. V59502) for increasing the stability of the fused
CC   protein as compared to the human protein only.
CC   The invention relates to 186 novel genes and their fragments (nucleic
CC   acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which
CC   are useful for preventing, treating or ameliorating  medical conditions
CC   e.g. by protein or gene therapy.  Also, pathological conditions can be
CC   diagnosed by determining the amount of the new polypeptides in a sample
CC   or by determining the presence of mutations in the new polynucleotides.
CC   Specific uses are described for each of the 186 polynucleotides, based on
CC   which tissues they are most highly expressed in (see V59511 for described
CC   uses).
XX
SQ   Sequence 1256 BP; 360 A; 283 C; 303 G; 309 T; 1 other;

  Query Match           51.0%;  Score 21.4;  DB 19;  Length 1256;
  Best Local Similarity   71.8%;  Pred. No. 1.1e+02;
  Matches   28;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      4 tctttgctcaaaatttgatcaaattttgagcaaagaata 42
          ||  |||||  || ||| |||||| |  ||||||| | |
Db    596 tcggtgctcggaacttgctcaaatctatagcaaagcaga 634


RESULT  18
V59618/c
ID   V59618 standard; DNA; 1256 BP.
XX
AC   V59618;
XX
DT   19-JAN-1999  (first entry)
XX
DE   Human secreted protein gene 108 clone HEBEK93.
XX
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS   Homo sapiens.
XX
PN   WO9839448-A2.
XX
PD   11-SEP-1998.
XX
PF   06-MAR-1998;   98WO-US04493.
XX
PR   02-OCT-1997;   97US-0061060.
PR   07-MAR-1997;   97US-0038621.
PR   07-MAR-1997;   97US-0040161.
PR   07-MAR-1997;   97US-0040162.
PR   07-MAR-1997;   97US-0040163.
PR   07-MAR-1997;   97US-0040333.
PR   07-MAR-1997;   97US-0040334.
PR   07-MAR-1997;   97US-0040336.
PR   07-MAR-1997;   97US-0040626.
PR   11-APR-1997;   97US-0043311.
PR   11-APR-1997;   97US-0043312.
PR   11-APR-1997;   97US-0043313.
PR   11-APR-1997;   97US-0043314.
PR   11-APR-1997;   97US-0043568.
PR   11-APR-1997;   97US-0043569.
PR   11-APR-1997;   97US-0043576.
PR   11-APR-1997;   97US-0043578.
PR   11-APR-1997;   97US-0043580.
PR   11-APR-1997;   97US-0043669.
PR   11-APR-1997;   97US-0043670.
PR   11-APR-1997;   97US-0043671.
PR   11-APR-1997;   97US-0043672.
PR   11-APR-1997;   97US-0043674.
PR   23-MAY-1997;   97US-0047492.
PR   23-MAY-1997;   97US-0047500.
PR   23-MAY-1997;   97US-0047501.
PR   23-MAY-1997;   97US-0047502.
PR   23-MAY-1997;   97US-0047503.
PR   23-MAY-1997;   97US-0047581.
PR   23-MAY-1997;   97US-0047582.
PR   23-MAY-1997;   97US-0047583.
PR   23-MAY-1997;   97US-0047584.
PR   23-MAY-1997;   97US-0047585.
PR   23-MAY-1997;   97US-0047586.
PR   23-MAY-1997;   97US-0047587.
PR   23-MAY-1997;   97US-0047588.
PR   23-MAY-1997;   97US-0047589.
```

PR 23-MAY-1997; 97US-0047590.
PR 23-MAY-1997; 97US-0047592.
PR 23-MAY-1997; 97US-0047593.
PR 23-MAY-1997; 97US-0047594.
PR 23-MAY-1997; 97US-0047595.
PR 23-MAY-1997; 97US-0047596.
PR 23-MAY-1997; 97US-0047597.
PR 23-MAY-1997; 97US-0047598.
PR 23-MAY-1997; 97US-0047599.
PR 23-MAY-1997; 97US-0047600.
PR 23-MAY-1997; 97US-0047601.
PR 23-MAY-1997; 97US-0047612.
PR 23-MAY-1997; 97US-0047613.
PR 23-MAY-1997; 97US-0047614.
PR 23-MAY-1997; 97US-0047615.
PR 23-MAY-1997; 97US-0047617.
PR 23-MAY-1997; 97US-0047618.
PR 23-MAY-1997; 97US-0047632.
PR 23-MAY-1997; 97US-0047633.
PR 06-JUN-1997; 97US-0048964.
PR 06-JUN-1997; 97US-0048974.
PR 13-JUN-1997; 97US-0049610.
PR 08-JUL-1997; 97US-0051926.
PR 16-JUL-1997; 97US-0052874.
PR 18-AUG-1997; 97US-0055724.
PR 22-AUG-1997; 97US-0056630.
PR 22-AUG-1997; 97US-0056631.
PR 22-AUG-1997; 97US-0056632.
PR 22-AUG-1997; 97US-0056636.
PR 22-AUG-1997; 97US-0056637.
PR 22-AUG-1997; 97US-0056662.
PR 22-AUG-1997; 97US-0056664.
PR 22-AUG-1997; 97US-0056845.
PR 22-AUG-1997; 97US-0056862.
PR 22-AUG-1997; 97US-0056864.
PR 22-AUG-1997; 97US-0056872.
PR 22-AUG-1997; 97US-0056874.
PR 22-AUG-1997; 97US-0056875.
PR 22-AUG-1997; 97US-0056876.
PR 22-AUG-1997; 97US-0056877.
PR 22-AUG-1997; 97US-0056878.
PR 22-AUG-1997; 97US-0056879.
PR 22-AUG-1997; 97US-0056880.
PR 22-AUG-1997; 97US-0056881.
PR 22-AUG-1997; 97US-0056882.
PR 22-AUG-1997; 97US-0056884.
PR 22-AUG-1997; 97US-0056886.
PR 22-AUG-1997; 97US-0056887.
PR 22-AUG-1997; 97US-0056888.
PR 22-AUG-1997; 97US-0056889.
PR 22-AUG-1997; 97US-0056892.
PR 22-AUG-1997; 97US-0056893.
PR 22-AUG-1997; 97US-0056894.
PR 22-AUG-1997; 97US-0056903.
PR 22-AUG-1997; 97US-0056908.
PR 22-AUG-1997; 97US-0056909.
PR 22-AUG-1997; 97US-0056910.
PR 22-AUG-1997; 97US-0056911.
PR 05-SEP-1997; 97US-0057650.
PR 05-SEP-1997; 97US-0057669.
PR 05-SEP-1997; 97US-0057761.
PR 12-SEP-1997; 97US-0058785.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;
PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
XX
DR WPI; 1998-506364/43.
DR P-PSDB; W74836.
XX
PT New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
PS Claim 1; Page 340-341; 721pp; English.
XX
CC This sequence represents a nucleic acid molecule designated Gene 108 from
CC the human cDNA clone HEBEK93 (deposited as clone ATCC 97901 and ATCC
CC 209047) which encodes a secreted human protein. The gene can be used to
CC generate fusion proteins by linking to the gene to a human immunoglobulin
CC Fc portion (e.g. V59502) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: V59511-V59812; amino acid sequences W74731-W75026) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 186 polynucleotides, based on
CC which tissues they are most highly expressed in (see V59511 for described
CC uses).
XX
SQ Sequence 1256 BP; 360 A; 283 C; 303 G; 309 T; 1 other;

Query Match 51.0%; Score 21.4; DB 19; Length 1256;
Best Local Similarity 71.8%; Pred. No. 1.1e+02;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 tattctttgctcaaaatttgatcaaattttgagcaaaga 39
| | ||||||| | |||||| ||| || ||||| ||
Db 634 TCTGCTTTGCTATAGATTTGAGCAAGTTCCGAGCACCGA 596

RESULT 19
Z35927
ID Z35927 standard; DNA; 1455 BP.
XX
AC Z35927;
XX
DT 07-FEB-2000 (first entry)
XX
DE Streptococcus pneumoniae pbp2B TER isolate h) nucleotide sequence.
XX
KW Streptococcus pneumoniae; penicillin binding protein; pbp2B; pbp1A;
KW transpeptidase encoding region; TER; antibiotic resistance; diagnosis;
KW detection; identification; pneumococcal meningitis; ss.
XX
OS Streptococcus pneumoniae.
XX
PN ZA9807024-A.
XX
PD 28-APR-1999.
XX
PF 05-AUG-1998; 98ZA-0007024.
XX
PR 01-AUG-1997; 97ZA-0006886.
XX
PA (SAME-) SOUTH AFRICAN INST MEDICAL RES.
PA (UYWI-) UNIV WITWATERSRAND.
PA (MEDI-) MEDICAL RES COUNCIL.
XX
PI Klugman KP, Smith AM, Du Plessis M;
XX
DR WPI; 1999-601770/51.
XX
PT Polymerase chain reaction assays for detecting Streptococcus pneumonia
PT useful for the diagnosis of pneumococcal meningitis -
XX
PS Claim 1; Fig 1; 63pp; English.
XX
CC A polymerase chain reaction (PCR) assays have been developed for
CC detecting an antibiotic resistant strain of Streptococcus pneumoniae

```
CC   using primers based on the penicillin binding protein 2B (pbp2B) gene
CC   and the pbp1A gene. The products and methods can be used for detecting
CC   S. pneumoniae, particularly antibiotic-resistant strains. They can be
CC   used for simultaneously diagnosing pneumococcal meningitis and
CC   identifying any antibiotic-resistant S. pneumoniae strains in a sample.
CC   The methods can be used for detecting S. pneumoniae strains resistant
CC   to antibiotics, e.g. beta-lactam antibiotics, preferably penicillin.
CC   The assays can be adapted to detect other pathogens causing meningitis.
CC   The assays can be used to detect an antibiotic resistant strain of
CC   S. pneumoniae with a minimum inhibitory concentration (MIC) of
CC   0.25-1 micro g/ml where the PCR products detected are a 1043 bp and
CC   a 224 bp product. The present sequence represents a Streptococcus
CC   pneumoniae pbp2B transpeptidase encoding region (TER) isolate nucleotide
CC   sequence from the present invention.
XX
SQ   Sequence 1455 BP; 424 A; 289 C; 331 G; 411 T; 0 other;


  Query Match             51.0%;  Score 21.4;  DB 20;  Length 1455;
  Best Local Similarity   71.8%;  Pred. No. 1.1e+02;
  Matches   28;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 attctttgctcaaaatttgatcaaattttgagcaaagaa 40
           || ||||||| |||| ||  |  ||||  |||| | |||
Db     355 atgctttgctgaaaagttatttcaattccgagctaggaa 393


RESULT  20
Z35927/c
ID   Z35927 standard; DNA; 1455 BP.
XX
AC   Z35927;
XX
DT   07-FEB-2000  (first entry)
XX
DE   Streptococcus pneumoniae pbp2B TER isolate h) nucleotide sequence.
XX
KW   Streptococcus pneumoniae; penicillin binding protein; pbp2B; pbp1A;
KW   transpeptidase encoding region; TER; antibiotic resistance; diagnosis;
KW   detection; identification; pneumococcal meningitis; ss.
XX
OS   Streptococcus pneumoniae.
XX
PN   ZA9807024-A.
XX
PD   28-APR-1999.
XX
PF   05-AUG-1998;   98ZA-0007024.
XX
PR   01-AUG-1997;   97ZA-0006886.
XX
PA   (SAME-) SOUTH AFRICAN INST MEDICAL RES.
PA   (UYWI-) UNIV WITWATERSRAND.
PA   (MEDI-) MEDICAL RES COUNCIL.
XX
PI   Klugman KP,  Smith AM,  Du Plessis M;
XX
DR   WPI; 1999-601770/51.
XX
PT   Polymerase chain reaction assays for detecting Streptococcus pneumonia
PT   useful for the diagnosis of pneumococcal meningitis   -
XX
PS   Claim 1; Fig 1; 63pp; English.
XX
CC   A polymerase chain reaction (PCR) assays have been developed for
CC   detecting an antibiotic resistant strain of Streptococcus pneumoniae
CC   using primers based on the penicillin binding protein 2B (pbp2B) gene
CC   and the pbp1A gene. The products and methods can be used for detecting
CC   S. pneumoniae, particularly antibiotic-resistant strains. They can be
CC   used for simultaneously diagnosing pneumococcal meningitis and
CC   identifying any antibiotic-resistant S. pneumoniae strains in a sample.
CC   The methods can be used for detecting S. pneumoniae strains resistant
CC   to antibiotics, e.g. beta-lactam antibiotics, preferably penicillin.
CC   The assays can be adapted to detect other pathogens causing meningitis.
CC   The assays can be used to detect an antibiotic resistant strain of
CC   S. pneumoniae with a minimum inhibitory concentration (MIC) of
CC   0.25-1 micro g/ml where the PCR products detected are a 1043 bp and
CC   a 224 bp product. The present sequence represents a Streptococcus
CC   pneumoniae pbp2B transpeptidase encoding region (TER) isolate nucleotide
CC   sequence from the present invention.
XX
SQ   Sequence 1455 BP; 424 A; 289 C; 331 G; 411 T; 0 other;


  Query Match             51.0%;  Score 21.4;  DB 20;  Length 1455;
  Best Local Similarity   71.8%;  Pred. No. 1.1e+02;
  Matches   28;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       3 ttctttgctcaaaatttgatcaaattttgagcaaagaat 41
           ||| | ||||  ||||  |  || |||| |||||||·||
Db     393 TTCCTAGCTCGGAATTGAAATAACTTTTCAGCAAAGCAT 355


Search completed: June  7, 2001, 18:09:57
Job time: 3455 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 7, 2001, 18:06:02 ; Search time 112.59 Seconds
(without alignments)
65.133 Million cell updates/sec

Title: US-09-601-561-2
Perfect score: 42
Sequence: 1 tattctttgctcaaaatttg.....caaattttgagcaaagaata 42

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	21.4	51.0	829	2	US-08-933-750C-87	Sequence 87, Appl
c	2	21.4	51.0	829	2	US-08-933-750C-87	Sequence 87, Appl
	3	21.4	51.0	829	4	US-09-234-613-87	Sequence 87, Appl
c	4	21.4	51.0	829	4	US-09-234-613-87	Sequence 87, Appl
	5	20.4	48.6	106	1	US-08-480-525-35	Sequence 35, Appl
c	6	20.4	48.6	106	1	US-08-480-525-35	Sequence 35, Appl
	7	20.4	48.6	106	5	PCT-US94-06422-35	Sequence 35, Appl
c	8	20.4	48.6	106	5	PCT-US94-06422-35	Sequence 35, Appl
	9	20.4	48.6	1419	1	US-08-480-525-1	Sequence 1, Appli
c	10	20.4	48.6	1419	1	US-08-480-525-1	Sequence 1, Appli
	11	20.4	48.6	1419	5	PCT-US94-06422-1	Sequence 1, Appli
c	12	20.4	48.6	1419	5	PCT-US94-06422-1	Sequence 1, Appli
	13	20.2	48.1	1183	1	US-08-188-582-25	Sequence 25, Appl
c	14	20.2	48.1	1183	1	US-08-188-582-25	Sequence 25, Appl
	15	20.2	48.1	1183	1	US-08-646-715-25	Sequence 25, Appl
c	16	20.2	48.1	1183	1	US-08-646-715-25	Sequence 25, Appl
	17	20.2	48.1	2038	1	US-08-181-271A-1	Sequence 1, Appli
c	18	20.2	48.1	2038	1	US-08-181-271A-1	Sequence 1, Appli
	19	20.2	48.1	2038	1	US-08-449-315-1	Sequence 1, Appli
c	20	20.2	48.1	2038	1	US-08-449-315-1	Sequence 1, Appli
	21	20.2	48.1	2038	1	US-08-444-803-1	Sequence 1, Appli
c	22	20.2	48.1	2038	1	US-08-444-803-1	Sequence 1, Appli
	23	20.2	48.1	2038	1	US-08-449-043-1	Sequence 1, Appli
c	24	20.2	48.1	2038	1	US-08-449-043-1	Sequence 1, Appli
	25	20.2	48.1	2038	1	US-08-456-265A-1	Sequence 1, Appli
c	26	20.2	48.1	2038	1	US-08-456-265A-1	Sequence 1, Appli
	27	20.2	48.1	2038	1	US-08-455-416-1	Sequence 1, Appli
c	28	20.2	48.1	2038	1	US-08-455-416-1	Sequence 1, Appli
	29	20.2	48.1	2038	1	US-08-455-244-1	Sequence 1, Appli
c	30	20.2	48.1	2038	1	US-08-455-244-1	Sequence 1, Appli
	31	20.2	48.1	2038	1	US-08-454-876-1	Sequence 1, Appli
c	32	20.2	48.1	2038	1	US-08-454-876-1	Sequence 1, Appli
	33	20.2	48.1	2038	2	US-08-457-364-1	Sequence 1, Appli
c	34	20.2	48.1	2038	2	US-08-457-364-1	Sequence 1, Appli
	35	20.2	48.1	2038	2	US-08-456-262-1	Sequence 1, Appli
c	36	20.2	48.1	2038	2	US-08-456-262-1	Sequence 1, Appli
	37	20.2	48.1	2038	2	US-08-456-240-1	Sequence 1, Appli
c	38	20.2	48.1	2038	2	US-08-456-240-1	Sequence 1, Appli
	39	20.2	48.1	2038	2	US-08-455-736-1	Sequence 1, Appli
c	40	20.2	48.1	2038	2	US-08-455-736-1	Sequence 1, Appli
	41	20.2	48.1	2038	2	US-08-971-217-1	Sequence 1, Appli
c	42	20.2	48.1	2038	2	US-08-971-217-1	Sequence 1, Appli
	43	19.8	47.1	1298	1	US-08-473-157A-1	Sequence 1, Appli
c	44	19.8	47.1	1298	1	US-08-473-157A-1	Sequence 1, Appli
	45	19.8	47.1	1542	2	US-08-865-311-1	Sequence 1, Appli
c	46	19.8	47.1	1542	2	US-08-865-311-1	Sequence 1, Appli
	47	19.8	47.1	2034	2	US-08-559-492-2	Sequence 2, Appli
c	48	19.8	47.1	2034	2	US-08-559-492-2	Sequence 2, Appli
	49	19.8	47.1	2635	3	US-08-714-918-38	Sequence 38, Appl
c	50	19.8	47.1	2635	3	US-08-714-918-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1
US-08-933-750C-87
; Sequence 87, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750C
; FILING DATE: September 23, 1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 87:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 829 base pairs

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  BRAITUT02
;      CLONE:  2098087
US-08-933-750C-87


  Query Match             51.0%;  Score 21.4;  DB 2;  Length 829;
  Best Local Similarity   71.8%;  Pred. No. 22;
  Matches   28;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      4 tctttgctcaaaatttgatcaaattttgagcaaagaata 42
          ||  |||||  || ||| |||||| |  ||||||| | |
Db    344 TCGGTGCTCGGAACTTGCTCAAATCTATAGCAAAGCAGA 382


RESULT   2
US-08-933-750C-87/c
; Sequence 87, Application US/08933750C
; Patent No. 5932442
;  GENERAL INFORMATION:
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Yue, Henry
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  HUMAN REGULATORY MOLECULES
;    NUMBER OF SEQUENCES:  98
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/933,750C
;      FILING DATE:  September 23, 1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0356 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  87:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  829 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  BRAITUT02
;      CLONE:  2098087
US-08-933-750C-87


  Query Match             51.0%;  Score 21.4;  DB 2;  Length 829;
  Best Local Similarity   71.8%;  Pred. No. 22;
  Matches   28;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaaga 39
          | | |||||||  | |||||| ||| ||  |||||  ||
Db    382 TCTGCTTTGCTATAGATTTGAGCAAGTTCCGAGCACCGA 344


RESULT   3
US-09-234-613-87
; Sequence 87, Application US/09234613
; Patent No. 6132973
;  GENERAL INFORMATION:
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Yue, Henry
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  HUMAN REGULATORY MOLECULES
;    NUMBER OF SEQUENCES:  98
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/234,613
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/933,750
;      FILING DATE:  September 23, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0356 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  87:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  829 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  BRAITUT02
;      CLONE:  2098087
US-09-234-613-87


  Query Match             51.0%;  Score 21.4;  DB 4;  Length 829;
  Best Local Similarity   71.8%;  Pred. No. 22;
  Matches   28;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      4 tctttgctcaaaatttgatcaaattttgagcaaagaata 42
          ||  |||||  || ||| |||||| |  ||||||| | |
Db    344 TCGGTGCTCGGAACTTGCTCAAATCTATAGCAAAGCAGA 382


RESULT   4
US-09-234-613-87/c
```

```
; Sequence 87, Application US/09234613
; Patent No. 6132973
;  GENERAL INFORMATION:
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Yue, Henry
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Corley, Neil C.
;    TITLE OF INVENTION:  HUMAN REGULATORY MOLECULES
;    NUMBER OF SEQUENCES:  98
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/234,613
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/933,750
;      FILING DATE:  September 23, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0356 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  87:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  829 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  BRAITUT02
;      CLONE:  2098087
US-09-234-613-87


  Query Match             51.0%;  Score 21.4;  DB 4;  Length 829;
  Best Local Similarity   71.8%;  Pred. No. 22;
  Matches   28;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaaga 39
          | | |||||||  | |||||| ||| ||  |||||  ||
Db    382 TCTGCTTTGCTATAGATTTGAGCAAGTTCCGAGCACCGA 344


RESULT   5
US-08-480-525-35
; Sequence 35, Application US/08480525
; Patent No. 5556754
;  GENERAL INFORMATION:
;    APPLICANT:  SINGER, DINAH, S.; KOHN,
;    APPLICANT:  LEONARD; MOZES, EDNA; SAJI, MOTOYASU;
;    APPLICANT:  WEISSMAN, JOCELYN; NAPOLITANO, GIORGIO;
;    APPLICANT:  LEDLEY, FRED, D
;    TITLE OF INVENTION:  METHODS OF TREATING
;    TITLE OF INVENTION:  AUTOIMMUNE DISEASES AND TRANSPLANTATION
;    TITLE OF INVENTION:  REJECTION
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,525
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/073,830
;      FILING DATE:  07-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAROL M. GRUPPI
;      REGISTRATION NUMBER:  P-37,341
;      REFERENCE/DOCKET NUMBER:  2026-4066
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  106
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  UNKNOWN
US-08-480-525-35


  Query Match             48.6%;  Score 20.4;  DB 1;  Length 106;
  Best Local Similarity   71.1%;  Pred. No. 40;
  Matches   27;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      4 tctttgctcaaaatttgatcaaattttgagcaaagaat 41
          | | | || |||||||||| ||||||| | | |  | |
Db     25 TATCTTCTAAAAATTTGATGAAATTTTAAACTATTATT 62


RESULT   6
US-08-480-525-35/c
; Sequence 35, Application US/08480525
; Patent No. 5556754
;  GENERAL INFORMATION:
;    APPLICANT:  SINGER, DINAH, S.; KOHN,
;    APPLICANT:  LEONARD; MOZES, EDNA; SAJI, MOTOYASU;
;    APPLICANT:  WEISSMAN, JOCELYN; NAPOLITANO, GIORGIO;
;    APPLICANT:  LEDLEY, FRED, D
;    TITLE OF INVENTION:  METHODS OF TREATING
;    TITLE OF INVENTION:  AUTOIMMUNE DISEASES AND TRANSPLANTATION
;    TITLE OF INVENTION:  REJECTION
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,525
;      FILING DATE:
```

```
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/073,830
;       FILING DATE:  07-JUN-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  CAROL M. GRUPPI
;       REGISTRATION NUMBER:  P-37,341
;       REFERENCE/DOCKET NUMBER:  2026-4066
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 758-4800
;       TELEFAX:  (212) 751-6849
;   INFORMATION FOR SEQ ID NO:  35:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  106
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  DOUBLE
;       TOPOLOGY:  UNKNOWN
US-08-480-525-35


  Query Match             48.6%;  Score 20.4;  DB 1;  Length 106;
  Best Local Similarity   71.1%;  Pred. No. 40;
  Matches   27;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 attctttgctcaaaatttgatcaaattttgagcaaaga 39
           | |   | | | ||||||| |||||||||| || | | |
Db      62 AATAATAGTTTAAAATTTCATCAAATTTTTAGAAGATA 25


RESULT   7
PCT-US94-06422-35
; Sequence 35, Application PC/TUS9406422
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  METHODS OF TREATING
;    TITLE OF INVENTION:  AUTOIMMUNE DISEASES AND TRANSPLANTATION
;    TITLE OF INVENTION:  REJECTION
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/06422
;      FILING DATE:  JUNE 7, 1994
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/073,830
;      FILING DATE:  07-JUN-1993
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAROL M. GRUPPI
;      REGISTRATION NUMBER:  37,341
;      REFERENCE/DOCKET NUMBER:  2026-4066PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  106
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  UNKNOWN
PCT-US94-06422-35


  Query Match             48.6%;  Score 20.4;  DB 5;  Length 106;
  Best Local Similarity   71.1%;  Pred. No. 40;
  Matches   27;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       4 tctttgctcaaaatttgatcaaattttgagcaaagaat 41
           | | | || |||||||||| ||||||| | | |  | |
Db      25 TATCTTCTAAAAATTTGATGAAATTTTAAACTATTATT 62


RESULT   8
PCT-US94-06422-35/c
; Sequence 35, Application PC/TUS9406422
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  METHODS OF TREATING
;    TITLE OF INVENTION:  AUTOIMMUNE DISEASES AND TRANSPLANTATION
;    TITLE OF INVENTION:  REJECTION
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/06422
;      FILING DATE:  JUNE 7, 1994
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/073,830
;      FILING DATE:  07-JUN-1993
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAROL M. GRUPPI
;      REGISTRATION NUMBER:  37,341
;      REFERENCE/DOCKET NUMBER:  2026-4066PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  106
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  UNKNOWN
PCT-US94-06422-35


  Query Match             48.6%;  Score 20.4;  DB 5;  Length 106;
  Best Local Similarity   71.1%;  Pred. No. 40;
  Matches   27;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 attctttgctcaaaatttgatcaaattttgagcaaaga 39
           | |   | | | ||||||| |||||||||| || | | |
Db      62 AATAATAGTTTAAAATTTCATCAAATTTTTAGAAGATA 25


RESULT   9
US-08-480-525-1
; Sequence 1, Application US/08480525
; Patent No. 5556754
;  GENERAL INFORMATION:
;    APPLICANT:  SINGER, DINAH, S.; KOHN,
;    APPLICANT:  LEONARD; MOZES, EDNA; SAJI, MOTOYASU;
;    APPLICANT:  WEISSMAN, JOCELYN; NAPOLITANO, GIORGIO;
;    APPLICANT:  LEDLEY, FRED, D
```

```
;    TITLE OF INVENTION:  METHODS OF TREATING
;    TITLE OF INVENTION:  AUTOIMMUNE DISEASES AND TRANSPLANTATION
;    TITLE OF INVENTION:  REJECTION
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,525
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/073,830
;      FILING DATE:  07-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAROL M. GRUPPI
;      REGISTRATION NUMBER:  P-37,341
;      REFERENCE/DOCKET NUMBER:  2026-4066
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1419
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  UNKNOWN
US-08-480-525-1


  Query Match             48.6%;  Score 20.4;  DB 1;  Length 1419;
  Best Local Similarity   71.1%;  Pred. No. 52;
  Matches   27;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      4 tctttgctcaaaatttgatcaaattttgagcaaagaat 41
          | | | || |||||||||| ||||||| | | |  | |
Db    612 TATCTTCTAAAAATTTGATGAAATTTTAAACTATTATT 649


RESULT  10
US-08-480-525-1/c
; Sequence 1, Application US/08480525
; Patent No. 5556754
;  GENERAL INFORMATION:
;    APPLICANT:  SINGER, DINAH, S.; KOHN,
;    APPLICANT:  LEONARD; MOZES, EDNA; SAJI, MOTOYASU;
;    APPLICANT:  WEISSMAN, JOCELYN; NAPOLITANO, GIORGIO;
;    APPLICANT:  LEDLEY, FRED, D
;    TITLE OF INVENTION:  METHODS OF TREATING
;    TITLE OF INVENTION:  AUTOIMMUNE DISEASES AND TRANSPLANTATION
;    TITLE OF INVENTION:  REJECTION
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,525
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/073,830
;      FILING DATE:  07-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAROL M. GRUPPI
;      REGISTRATION NUMBER:  P-37,341
;      REFERENCE/DOCKET NUMBER:  2026-4066
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1419
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  UNKNOWN
US-08-480-525-1


  Query Match             48.6%;  Score 20.4;  DB 1;  Length 1419;
  Best Local Similarity   71.1%;  Pred. No. 52;
  Matches   27;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 attctttgctcaaaatttgatcaaattttgagcaaaga 39
            | |  | | | ||||||| |||||||||| || | | |
Db    649 AATAATAGTTTAAAATTTCATCAAATTTTTAGAAGATA 612


RESULT  11
PCT-US94-06422-1
; Sequence 1, Application PC/TUS9406422
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  METHODS OF TREATING
;    TITLE OF INVENTION:  AUTOIMMUNE DISEASES AND TRANSPLANTATION
;    TITLE OF INVENTION:  REJECTION
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/06422
;      FILING DATE:  JUNE 7, 1994
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/073,830
;      FILING DATE:  07-JUN-1993
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAROL M. GRUPPI
;      REGISTRATION NUMBER:  37,341
;      REFERENCE/DOCKET NUMBER:  2026-4066PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1419
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
```

```
;      TOPOLOGY:  UNKNOWN
PCT-US94-06422-1


  Query Match             48.6%;  Score 20.4;  DB 5;  Length 1419;
  Best Local Similarity   71.1%;  Pred. No. 52;
  Matches   27;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      4 tctttgctcaaaatttgatcaaattttgagcaaagaat 41
          | | | || |||||||||| ||||||| | | |  | |
Db    612 TATCTTCTAAAAATTTGATGAAATTTTAAACTATTATT 649


RESULT  12
PCT-US94-06422-1/c
; Sequence 1, Application PC/TUS9406422
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  METHODS OF TREATING
;    TITLE OF INVENTION:  AUTOIMMUNE DISEASES AND TRANSPLANTATION
;    TITLE OF INVENTION:  REJECTION
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/06422
;      FILING DATE:  JUNE 7, 1994
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/073,830
;      FILING DATE:  07-JUN-1993
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAROL M. GRUPPI
;      REGISTRATION NUMBER:  37,341
;      REFERENCE/DOCKET NUMBER:  2026-4066PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1419
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE
;      TOPOLOGY:  UNKNOWN
PCT-US94-06422-1


  Query Match             48.6%;  Score 20.4;  DB 5;  Length 1419;
  Best Local Similarity   71.1%;  Pred. No. 52;
  Matches   27;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 attctttgctcaaaatttgatcaaattttgagcaaaga 39
          | |  | | | ||||||| |||||||||| || | | |
Db    649 AATAATAGTTTAAAATTTCATCAAATTTTTAGAAGATA 612


RESULT  13
US-08-188-582-25
; Sequence 25, Application US/08188582
; Patent No. 5534410
;  GENERAL INFORMATION:
;    APPLICANT:  Tjian, Robert
;    APPLICANT:  Comai, Lucio
;    APPLICANT:  Dynlact, Brian D.
;    APPLICANT:  Hoey, Timothy
;    APPLICANT:  Ruppert, Siegfried
;    APPLICANT:  Tanese, Naoko
;    APPLICANT:  Wang, Edith
;    APPLICANT:  Weinzierl, Robert O.J.
;    TITLE OF INVENTION:  TATA-BINDING PROTEIN ASSOCIATED FACTORS,
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/188,582
;      FILING DATE:  28-JAN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-57650-2/AJT/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1183 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  161..952
US-08-188-582-25


  Query Match             48.1%;  Score 20.2;  DB 1;  Length 1183;
  Best Local Similarity   68.3%;  Pred. No. 60;
  Matches   28;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      2 attctttgctcaaaatttgatcaaattttgagcaaagaata 42
          | ||||| |  ||||    ||||| || || |  |||||||
Db    549 AATCTTTACAGAAAAAGGCATCAACTTCTGCGGGAAGAATA 589


RESULT  14
US-08-188-582-25/c
; Sequence 25, Application US/08188582
; Patent No. 5534410
;  GENERAL INFORMATION:
;    APPLICANT:  Tjian, Robert
;    APPLICANT:  Comai, Lucio
;    APPLICANT:  Dynlact, Brian D.
;    APPLICANT:  Hoey, Timothy
;    APPLICANT:  Ruppert, Siegfried
;    APPLICANT:  Tanese, Naoko
;    APPLICANT:  Wang, Edith
;    APPLICANT:  Weinzierl, Robert O.J.
;    TITLE OF INVENTION:  TATA-BINDING PROTEIN ASSOCIATED FACTORS,
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;       STREET:  4 Embarcadero Center, Suite 3400
;       CITY:  San Francisco
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  94111-4187
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/188,582
;       FILING DATE:  28-JAN-1994
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Osman, Richard A
;       REGISTRATION NUMBER:  36,627
;       REFERENCE/DOCKET NUMBER:  A-57650-2/AJT/RAO
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 781-1989
;       TELEFAX:  (415) 398-3249
;       TELEX:  910 277299
;   INFORMATION FOR SEQ ID NO:  25:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1183 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  161..952
US-08-188-582-25


  Query Match             48.1%;  Score 20.2;  DB 1;  Length 1183;
  Best Local Similarity   68.3%;  Pred. No. 60;
  Matches   28;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaagaat 41
          |||||||  | || || |||||    ||||  | |||||
Db    589 TATTCTTCCCGCAGAAGTTGATGCCTTTTTCTGTAAAGATT 549


RESULT  15
US-08-646-715-25
; Sequence 25, Application US/08646715
; Patent No. 5637686
;  GENERAL INFORMATION:
;    APPLICANT:  Tjian, Robert
;    APPLICANT:  Comai, Lucio
;    APPLICANT:  Dynlact, Brian D.
;    APPLICANT:  Hoey, Timothy
;    APPLICANT:  Ruppert, Siegfried
;    APPLICANT:  Tanese, Naoko
;    APPLICANT:  Wang, Edith
;    APPLICANT:  Weinzierl, Robert O.J.
;    TITLE OF INVENTION:  TATA-BINDING PROTEIN ASSOCIATED FACTORS,
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/646,715
;      FILING DATE:  09-MAY-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/188,582
;      FILING DATE:  28-JAN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-57650-2/AJT/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1183 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  161..952
US-08-646-715-25


  Query Match             48.1%;  Score 20.2;  DB 1;  Length 1183;
  Best Local Similarity   68.3%;  Pred. No. 60;
  Matches   28;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      2 attctttgctcaaaatttgatcaaattttgagcaaagaata 42
            | |||||  |  ||||    |||||  ||  ||  |  |||||||
Db    549 AATCTTTACAGAAAAAGGCATCAACTTCTGCGGGAAGAATA 589


RESULT  16
US-08-646-715-25/c
; Sequence 25, Application US/08646715
; Patent No. 5637686
;  GENERAL INFORMATION:
;    APPLICANT:  Tjian, Robert
;    APPLICANT:  Comai, Lucio
;    APPLICANT:  Dynlact, Brian D.
;    APPLICANT:  Hoey, Timothy
;    APPLICANT:  Ruppert, Siegfried
;    APPLICANT:  Tanese, Naoko
;    APPLICANT:  Wang, Edith
;    APPLICANT:  Weinzierl, Robert O.J.
;    TITLE OF INVENTION:  TATA-BINDING PROTEIN ASSOCIATED FACTORS,
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/646,715
;      FILING DATE:  09-MAY-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/188,582
;      FILING DATE:  28-JAN-1994
;    ATTORNEY/AGENT INFORMATION:
```

```
;       NAME:  Osman, Richard A
;       REGISTRATION NUMBER:  36,627
;       REFERENCE/DOCKET NUMBER:  A-57650-2/AJT/RAO
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 781-1989
;       TELEFAX:  (415) 398-3249
;       TELEX:  910 277299
;   INFORMATION FOR SEQ ID NO:  25:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1183 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  161..952
US-08-646-715-25


  Query Match             48.1%;  Score 20.2;  DB 1;  Length 1183;
  Best Local Similarity   68.3%;  Pred. No. 60;
  Matches   28;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       1 tattctttgctcaaaatttgatcaaattttgagcaaagaat 41
           |||||||  | || || |||||    ||||  | ||||| |
Db     589 TATTCTTCCCGCAGAAGTTGATGCCTTTTTCTGTAAAGATT 549


RESULT  17
US-08-181-271A-1
; Sequence 1, Application US/08181271A
; Patent No. 5614395
;  GENERAL INFORMATION:
;     APPLICANT:  Ryals, John A.
;     APPLICANT:  Alexander, Danny C.
;     APPLICANT:  Beck, James J.
;     APPLICANT:  Duesing, John H.
;     APPLICANT:  Friedrich, Leslie B.
;     APPLICANT:  Goodman, Robert M.
;     APPLICANT:  Harms, Christian
;     APPLICANT:  Meins, Jr., Frederick
;     APPLICANT:  Montoya, Alice
;     APPLICANT:  Moyer, Mary B.
;     APPLICANT:  Neuhaus, Jean-Marc
;     APPLICANT:  Payne, George B.
;     APPLICANT:  Sperison, Christoph
;     APPLICANT:  Stinson, Jeffrey R.
;     APPLICANT:  Uknes, Scott J.
;     APPLICANT:  Ward, Eric R.
;     APPLICANT:  Williams, Shericca C.
;     TITLE OF INVENTION:  CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
;     TITLE OF INVENTION:  DNA SEQUENCES AND USES THEREOF
;     NUMBER OF SEQUENCES:  106
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  CIBA-GEIGY Corporation
;       STREET:  7 Skyline Drive
;       CITY:  Hawthorne
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10532
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/181,271A
;       FILING DATE:  13-JAN-94
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/093,301
;       FILING DATE:  16-JUL-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/937,197
;       FILING DATE:  6-NOV-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/678,378
;       FILING DATE:  1-APR-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/305,566
;       FILING DATE:  6-FEB-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/165,667
;       FILING DATE:  8-MAR-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/042,847
;       FILING DATE:  6-APR-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/632,441
;       FILING DATE:  21-DEC-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/425,504
;       FILING DATE:  20-OCT 1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/848,506
;       FILING DATE:  6-MAR-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/768,122
;       FILING DATE:  27-SEP-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/580,431
;       FILING DATE:  7-SEP-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/368,672
;       FILING DATE:  20-JUN-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/329,018
;       FILING DATE:  24-MAR-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/045,957
;       FILING DATE:  12-APR-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Elmer, James Scott
;       REGISTRATION NUMBER:  36,129
;       REFERENCE/DOCKET NUMBER:  S-19825/P1/CGC 1727
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (919)541-8614
;       TELEFAX:  (919)541-8689
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2038 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  932..1435
US-08-181-271A-1


  Query Match             48.1%;  Score 20.2;  DB 1;  Length 2038;
  Best Local Similarity   75.8%;  Pred. No. 63;
  Matches   25;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       8 tgctcaaaatttgatcaaattttgagcaaagaa 40
             || ||||  |||   ||||| ||||| ||||||
Db     331 TGTTCAAGTTTTCCACAAATATTGAGAAAAGAA 363


RESULT  18
US-08-181-271A-1/c
; Sequence 1, Application US/08181271A
; Patent No. 5614395
;  GENERAL INFORMATION:
```

; APPLICANT: Ryals, John A.
; APPLICANT: Alexander, Danny C.
; APPLICANT: Beck, James J.
; APPLICANT: Duesing, John H.
; APPLICANT: Friedrich, Leslie B.
; APPLICANT: Goodman, Robert M.
; APPLICANT: Harms, Christian
; APPLICANT: Meins, Jr., Frederick
; APPLICANT: Montoya, Alice
; APPLICANT: Moyer, Mary B.
; APPLICANT: Neuhaus, Jean-Marc
; APPLICANT: Payne, George B.
; APPLICANT: Sperison, Christoph
; APPLICANT: Stinson, Jeffrey R.
; APPLICANT: Uknes, Scott J.
; APPLICANT: Ward, Eric R.
; APPLICANT: Williams, Shericca C.
; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/181,271A
; FILING DATE: 13-JAN-94
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/093,301
; FILING DATE: 16-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,197
; FILING DATE: 6-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/678,378
; FILING DATE: 1-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/305,566
; FILING DATE: 6-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/165,667
; FILING DATE: 8-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/042,847
; FILING DATE: 6-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/632,441
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/425,504
; FILING DATE: 20-OCT 1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,506
; FILING DATE: 6-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/768,122
; FILING DATE: 27-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/580,431
; FILING DATE: 7-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/368,672
; FILING DATE: 20-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/329,018
; FILING DATE: 24-MAR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/045,957
; FILING DATE: 12-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8614
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2038 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 932..1435
US-08-181-271A-1

Query Match 48.1%; Score 20.2; DB 1; Length 2038;
Best Local Similarity 75.8%; Pred. No. 63;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 3 ttctttgctcaaaatttgatcaaattttgagca 35
||||||| ||||| ||||| ||| |||| ||
Db 363 TTCTTTTCTCAATATTTGTGGAAAACTTGAACA 331

RESULT 19
US-08-449-315-1
; Sequence 1, Application US/08449315
; Patent No. 5650505
; GENERAL INFORMATION:
; APPLICANT: Ryals, John A.
; APPLICANT: Alexander, Danny C.
; APPLICANT: Beck, James J.
; APPLICANT: Duesing, John H.
; APPLICANT: Friedrich, Leslie B.
; APPLICANT: Goodman, Robert M.
; APPLICANT: Harms, Christian
; APPLICANT: Meins, Jr., Frederick
; APPLICANT: Montoya, Alice
; APPLICANT: Moyer, Mary B.
; APPLICANT: Neuhaus, Jean-Marc
; APPLICANT: Payne, George B.
; APPLICANT: Sperison, Christoph
; APPLICANT: Stinson, Jeffrey R.
; APPLICANT: Uknes, Scott J.
; APPLICANT: Ward, Eric R.
; APPLICANT: Williams, Shericca C.
; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,315

```
;       FILING DATE:  24-MAY-1995
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/181,271
;       FILING DATE:  13-JAN-94
;       APPLICATION NUMBER:  US 08/093,301
;       FILING DATE:  16-JUL-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/937,197
;       FILING DATE:  6-NOV-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/678,378
;       FILING DATE:  1-APR-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/305,566
;       FILING DATE:  6-FEB-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/165,667
;       FILING DATE:  8-MAR-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/042,847
;       FILING DATE:  6-APR-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/632,441
;       FILING DATE:  21-DEC-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/425,504
;       FILING DATE:  20-OCT 1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/848,506
;       FILING DATE:  6-MAR-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/768,122
;       FILING DATE:  27-SEP-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/580,431
;       FILING DATE:  7-SEP-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/368,672
;       FILING DATE:  20-JUN-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/329,018
;       FILING DATE:  24-MAR-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/045,957
;       FILING DATE:  12-APR-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Elmer, James Scott
;       REGISTRATION NUMBER:  36,129
;       REFERENCE/DOCKET NUMBER:  S-19825/P1/CGC 1727
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (919)541-8614
;       TELEFAX:  (919)541-8689
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2038 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  932..1435
US-08-449-315-1


  Query Match             48.1%;  Score 20.2;  DB 1;  Length 2038;
  Best Local Similarity   75.8%;  Pred. No. 63;
  Matches   25;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       8 tgctcaaaatttgatcaaattttgagcaaagaa 40
           || ||||  |||   ||||| ||||| ||||||
Db     331 TGTTCAAGTTTTCCACAAATATTGAGAAAAGAA 363
```

```
RESULT  20
US-08-449-315-1/c
; Sequence 1, Application US/08449315
; Patent No. 5650505
;  GENERAL INFORMATION:
;     APPLICANT:  Ryals, John A.
;     APPLICANT:  Alexander, Danny C.
;     APPLICANT:  Beck, James J.
;     APPLICANT:  Duesing, John H.
;     APPLICANT:  Friedrich, Leslie B.
;     APPLICANT:  Goodman, Robert M.
;     APPLICANT:  Harms, Christian
;     APPLICANT:  Meins, Jr., Frederick
;     APPLICANT:  Montoya, Alice
;     APPLICANT:  Moyer, Mary B.
;     APPLICANT:  Neuhaus, Jean-Marc
;     APPLICANT:  Payne, George B.
;     APPLICANT:  Sperison, Christoph
;     APPLICANT:  Stinson, Jeffrey R.
;     APPLICANT:  Uknes, Scott J.
;     APPLICANT:  Ward, Eric R.
;     APPLICANT:  Williams, Shericca C.
;     TITLE OF INVENTION:  CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
;     TITLE OF INVENTION:  DNA SEQUENCES AND USES THEREOF
;     NUMBER OF SEQUENCES:  106
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  CIBA-GEIGY Corporation
;       STREET:  7 Skyline Drive
;       CITY:  Hawthorne
;       STATE:  New York
;       COUNTRY:  USA
;       ZIP:  10532
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/449,315
;       FILING DATE:  24-MAY-1995
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/181,271
;       FILING DATE:  13-JAN-94
;       APPLICATION NUMBER:  US 08/093,301
;       FILING DATE:  16-JUL-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/937,197
;       FILING DATE:  6-NOV-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/678,378
;       FILING DATE:  1-APR-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/305,566
;       FILING DATE:  6-FEB-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/165,667
;       FILING DATE:  8-MAR-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/042,847
;       FILING DATE:  6-APR-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/632,441
;       FILING DATE:  21-DEC-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/425,504
;       FILING DATE:  20-OCT 1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/848,506
;       FILING DATE:  6-MAR-1992
;     PRIOR APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  US 07/768,122
;       FILING DATE:  27-SEP-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/580,431
;       FILING DATE:  7-SEP-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/368,672
;       FILING DATE:  20-JUN-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/329,018
;       FILING DATE:  24-MAR-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/045,957
;       FILING DATE:  12-APR-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Elmer, James Scott
;       REGISTRATION NUMBER:  36,129
;       REFERENCE/DOCKET NUMBER:  S-19825/P1/CGC 1727
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (919)541-8614
;       TELEFAX:  (919)541-8689
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2038 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  932..1435
US-08-449-315-1


  Query Match             48.1%;  Score 20.2;  DB 1;  Length 2038;
  Best Local Similarity   75.8%;  Pred. No. 63;
  Matches   25;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       3 ttctttgctcaaaatttgatcaaattttgagca 35
           |||||| ||||| |||||   |||  |||| ||
Db     363 TTCTTTTCTCAATATTTGTGGAAAACTTGAACA 331


Search completed: June  7, 2001, 18:06:03
Job time: 4421 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 7, 2001, 17:41:28 ; Search time 1600.48 Seconds
(without alignments)
229.254 Million cell updates/sec

Title: US-09-601-561-2
Perfect score: 42
Sequence: 1 tattctttgctcaaaatttg.....caaattttgagcaaagaata 42

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_gss_fun:*
177: em_gss_hum1:*
178: em_gss_hum2:*
179: em_gss_hum3:*
180: em_gss_hum4:*
181: em_gss_hum5:*
182: em_gss_hum6:*
183: em_gss_hum7:*
184: em_gss_hum8:*
185: em_gss_hum9:*
186: em_gss_inv1:*
187: em_gss_inv2:*
188: em_gss_inv3:*
189: em_gss_other:*
190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		25.6	61.0	281	166	BE420329	BE420329 WWS05.D5R
c 2		25.6	61.0	281	166	BE420329	BE420329 WWS05.D5R
3		25.6	61.0	511	166	BE420374	BE420374 WWS05.H7R
c 4		25.6	61.0	511	166	BE420374	BE420374 WWS05.H7R
5		24.8	59.0	1101	229	CNS00GIW	AL072245 Drosophil
c 6		24.8	59.0	1101	229	CNS00GIW	AL072245 Drosophil
7		24.4	58.1	502	211	AQ751027	AQ751027 HS_5576_B
c 8		24.4	58.1	502	211	AQ751027	AQ751027 HS_5576_B
9		24.4	58.1	714	32	AV718044	AV718044 AV718044
c 10		24.4	58.1	714	32	AV718044	AV718044 AV718044
11		24.4	58.1	811	205	AQ327650	AQ327650 nbxb0041B
c 12		24.4	58.1	811	205	AQ327650	AQ327650 nbxb0041B
13		24.2	57.6	344	153	H55810	H55810 yr02c11.r1
c 14		24.2	57.6	344	153	H55810	H55810 yr02c11.r1
15		24	57.1	456	143	BF052039	BF052039 EST437286
c 16		24	57.1	456	143	BF052039	BF052039 EST437286
17		24	57.1	509	173	BG135705	BG135705 EST468597
c 18		24	57.1	509	173	BG135705	BG135705 EST468597

	19	24	57.1	557	121	AW928999	AW928999 EST337883
c	20	24	57.1	557	121	AW928999	AW928999 EST337883
	21	23.8	56.7	299	124	BB088990	BB088990 BB088990
c	22	23.8	56.7	299	124	BB088990	BB088990 BB088990
	23	23.8	56.7	770	205	AQ330965	AQ330965 nbxb0048B
c	24	23.8	56.7	770	205	AQ330965	AQ330965 nbxb0048B
	25	23.4	55.7	99	113	AW285120	AW285120 LG1_300_H
c	26	23.4	55.7	99	113	AW285120	AW285120 LG1_300_H
	27	23.4	55.7	299	214	AZ001156	AZ001156 RPCI-23-3
c	28	23.4	55.7	299	214	AZ001156	AZ001156 RPCI-23-3
	29	23.4	55.7	577	205	AQ379041	AQ379041 RPCI-11-1
c	30	23.4	55.7	577	205	AQ379041	AQ379041 RPCI-11-1
	31	23.2	55.2	414	218	AZ227022	AZ227022 RPCI-23-6
c	32	23.2	55.2	414	218	AZ227022	AZ227022 RPCI-23-6
	33	23.2	55.2	990	140	BE859252	BE859252 SsS0896 S
c	34	23.2	55.2	990	140	BE859252	BE859252 SsS0896 S
	35	23	54.8	432	214	AQ964351	AQ964351 LERGV55TF
c	36	23	54.8	432	214	AQ964351	AQ964351 LERGV55TF
	37	23	54.8	493	209	AQ627847	AQ627847 CITBI-E1-
c	38	23	54.8	493	209	AQ627847	AQ627847 CITBI-E1-
	39	23	54.8	495	221	AZ443693	AZ443693 1M0238B09
c	40	23	54.8	495	221	AZ443693	AZ443693 1M0238B09
	41	23	54.8	501	121	AW856899	AW856899 RC0-CT030
c	42	23	54.8	501	121	AW856899	AW856899 RC0-CT030
	43	22.8	54.3	226	29	AV365142	AV365142 AV365142
c	44	22.8	54.3	226	29	AV365142	AV365142 AV365142
	45	22.8	54.3	236	29	AV365865	AV365865 AV365865
c	46	22.8	54.3	236	29	AV365865	AV365865 AV365865
	47	22.8	54.3	253	29	AV351557	AV351557 AV351557
c	48	22.8	54.3	253	29	AV351557	AV351557 AV351557
	49	22.8	54.3	304	128	BB247085	BB247085 BB247085
c	50	22.8	54.3	304	128	BB247085	BB247085 BB247085

ALIGNMENTS

```
RESULT   1
BE420329
LOCUS       BE420329      281 bp    mRNA            EST       24-JUL-2000
DEFINITION  WWS05.D5R000101 ITEC WWS Wheat Scutellum Library Triticum aestivum
            cDNA clone WWS05.D5, mRNA sequence.
ACCESSION   BE420329
VERSION     BE420329.1  GI:9418175
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae;
            Triticum.
REFERENCE   1  (bases 1 to 281)
  AUTHORS   Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
            ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
            Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
            Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
            Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
            Sorrells,M., Warburton,M. and Wenzel,G.
  TITLE     International Triticeae EST Cooperative (ITEC): Production of
            Expressed Sequence Tags for Species of the Triticeae
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Schuch W
            Zeneca Wheat Improvement Centre, Norwich Research Park
            Colney Lane, Norwich NR4 7UH UNITED KINGDOM
            Tel: 44 1603 250 2600
            Fax: 44 1603 250 699
            Email: wolfgang.schuch@aguk.zeneca.com
            International Triticeae EST Cooperative (ITEC)
            http://wheat.pw.usda.gov/genome.
FEATURES             Location/Qualifiers
     source          1. .281
                     /organism="Triticum aestivum"
                     /cultivar="Novosibirskaya 67"
                     /db_xref="taxon:4565"
                     /clone="WWS05.D5"
                     /clone_lib="ITEC WWS Wheat Scutellum Library"
                     /tissue_type="scutellum callus"
                     /note="M13 Reverse sequencing primer used for 5' end of
                     clone. "
BASE COUNT       77 a      52 c      67 g      85 t
ORIGIN


  Query Match             61.0%;  Score 25.6;  DB 166;  Length 281;
  Best Local Similarity   77.5%;  Pred. No. 1.1e+02;
  Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaagaa 40
          | ||||||| ||  |||||||||  ||  || ||||||||
Db    168 TTTTCTTTGGTCCTAATTTGATCTCATGGTGTGCAAAGAA 207


RESULT   2
BE420329/c
LOCUS       BE420329      281 bp    mRNA            EST       24-JUL-2000
DEFINITION  WWS05.D5R000101 ITEC WWS Wheat Scutellum Library Triticum aestivum
            cDNA clone WWS05.D5, mRNA sequence.
ACCESSION   BE420329
VERSION     BE420329.1  GI:9418175
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae;
            Triticum.
REFERENCE   1  (bases 1 to 281)
  AUTHORS   Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
            ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
            Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
            Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
            Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
            Sorrells,M., Warburton,M. and Wenzel,G.
  TITLE     International Triticeae EST Cooperative (ITEC): Production of
            Expressed Sequence Tags for Species of the Triticeae
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Schuch W
            Zeneca Wheat Improvement Centre, Norwich Research Park
            Colney Lane, Norwich NR4 7UH UNITED KINGDOM
            Tel: 44 1603 250 2600
            Fax: 44 1603 250 699
            Email: wolfgang.schuch@aguk.zeneca.com
            International Triticeae EST Cooperative (ITEC)
            http://wheat.pw.usda.gov/genome.
FEATURES             Location/Qualifiers
     source          1. .281
                     /organism="Triticum aestivum"
                     /cultivar="Novosibirskaya 67"
                     /db_xref="taxon:4565"
                     /clone="WWS05.D5"
                     /clone_lib="ITEC WWS Wheat Scutellum Library"
                     /tissue_type="scutellum callus"
                     /note="M13 Reverse sequencing primer used for 5' end of
                     clone. "
BASE COUNT       77 a      52 c      67 g      85 t
ORIGIN


  Query Match             61.0%;  Score 25.6;  DB 166;  Length 281;
  Best Local Similarity   77.5%;  Pred. No. 1.1e+02;
  Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      3 ttctttgctcaaaatttgatcaaattttgagcaaagaata 42
          |||||||| ||  ||  |||||||||  || ||||||| |
Db    207 TTCTTTGCACACCATGAGATCAAATTAGGACCAAAGAAAA 168


RESULT   3
```

```
BE420374
LOCUS       BE420374      511 bp    mRNA            EST       24-JUL-2000
DEFINITION  WWS05.H7R000101 ITEC WWS Wheat Scutellum Library Triticum aestivum
            cDNA clone WWS05.H7, mRNA sequence.
ACCESSION   BE420374
VERSION     BE420374.1  GI:9418220
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae;
            Triticum.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
            ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
            Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
            Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
            Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
            Sorrells,M., Warburton,M. and Wenzel,G.
  TITLE     International Triticeae EST Cooperative (ITEC): Production of
            Expressed Sequence Tags for Species of the Triticeae
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Schuch W
            Zeneca Wheat Improvement Centre, Norwich Research Park
            Colney Lane, Norwich NR4 7UH UNITED KINGDOM
            Tel: 44 1603 250 2600
            Fax: 44 1603 250 699
            Email: wolfgang.schuch@aguk.zeneca.com
            International Triticeae EST Cooperative (ITEC)
            http://wheat.pw.usda.gov/genome.
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Triticum aestivum"
                     /cultivar="Novosibirskaya 67"
                     /db_xref="taxon:4565"
                     /clone="WWS05.H7"
                     /clone_lib="ITEC WWS Wheat Scutellum Library"
                     /tissue_type="scutellum callus"
                     /note="M13 Reverse sequencing primer used for 5' end of
                     clone. "
BASE COUNT      144 a     88 c    118 g    161 t
ORIGIN


  Query Match             61.0%;  Score 25.6;  DB 166;  Length 511;
  Best Local Similarity   77.5%;  Pred. No. 1.1e+02;
  Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaagaa 40
          | ||||||| ||  |||||||||  ||  || ||||||||
Db    168 TTTTCTTTGGTCCTAATTTGATCTCATGGTGTGCAAAGAA 207


RESULT   4
BE420374/c
LOCUS       BE420374      511 bp    mRNA            EST       24-JUL-2000
DEFINITION  WWS05.H7R000101 ITEC WWS Wheat Scutellum Library Triticum aestivum
            cDNA clone WWS05.H7, mRNA sequence.
ACCESSION   BE420374
VERSION     BE420374.1  GI:9418220
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae;
            Triticum.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
            ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
            Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
            Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
            Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
            Sorrells,M., Warburton,M. and Wenzel,G.
  TITLE     International Triticeae EST Cooperative (ITEC): Production of
            Expressed Sequence Tags for Species of the Triticeae
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Schuch W
            Zeneca Wheat Improvement Centre, Norwich Research Park
            Colney Lane, Norwich NR4 7UH UNITED KINGDOM
            Tel: 44 1603 250 2600
            Fax: 44 1603 250 699
            Email: wolfgang.schuch@aguk.zeneca.com
            International Triticeae EST Cooperative (ITEC)
            http://wheat.pw.usda.gov/genome.
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Triticum aestivum"
                     /cultivar="Novosibirskaya 67"
                     /db_xref="taxon:4565"
                     /clone="WWS05.H7"
                     /clone_lib="ITEC WWS Wheat Scutellum Library"
                     /tissue_type="scutellum callus"
                     /note="M13 Reverse sequencing primer used for 5' end of
                     clone. "
BASE COUNT      144 a     88 c    118 g    161 t
ORIGIN


  Query Match             61.0%;  Score 25.6;  DB 166;  Length 511;
  Best Local Similarity   77.5%;  Pred. No. 1.1e+02;
  Matches   31;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      3 ttctttgctcaaaatttgatcaaattttgagcaaagaata 42
          |||||||| ||  ||  |||||||||  || ||||||| |
Db    207 TTCTTTGCACACCATGAGATCAAATTAGGACCAAAGAAAA 168


RESULT   5
CNS00GIW
LOCUS       CNS00GIW     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR33L06 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL072245
VERSION     AL072245.1  GI:4952025
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
```

```
                     /clone_lib="RPCI-98"
                     /clone="BACR33L06"
                     /note="end : T7"
BASE COUNT      335 a    187 c    121 g    329 t    129 others
ORIGIN


  Query Match             59.0%;  Score 24.8;  DB 229;  Length 1101;
  Best Local Similarity   67.5%;  Pred. No. 2.1e+02;
  Matches   27;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      3 ttctttgctcaaaatttgatcaaattttgagcaaagaata 42
          || |||  :|:|||||| || |||||||::  || :|||| 
Db    864 TTTTTTTTKCRAAATTTCATAAAATTTTSRTAAATDAATA 903


RESULT   6
CNS00GIW/c
LOCUS       CNS00GIW     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR33L06 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL072245
VERSION     AL072245.1  GI:4952025
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR33L06"
                     /note="end : T7"
BASE COUNT      335 a    187 c    121 g    329 t    129 others
ORIGIN


  Query Match             59.0%;  Score 24.8;  DB 229;  Length 1101;
  Best Local Similarity   67.5%;  Pred. No. 2.1e+02;
  Matches   27;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaagaa 40
          ||||: ||  ::||||||| || ||||||:|:  ||| ||
Db    903 TATTHATTTAYSAAAATTTTATGAAATTTYGMAAAAAAAA 864


RESULT   7
AQ751027
```

```
LOCUS       AQ751027      502 bp    DNA             GSS       19-JUL-1999
DEFINITION  HS_5576_B1_B09_SP6 RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1152 Col=17 Row=D, DNA sequence.
ACCESSION   AQ751027
VERSION     AQ751027.1  GI:5538185
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1152  row: D  column: 17
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 502.
FEATURES             Location/Qualifiers
     source          1. .502
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=1152 Col=17 Row=D"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      122 a    118 c     71 g    191 t
ORIGIN


  Query Match             58.1%;  Score 24.4;  DB 211;  Length 502;
  Best Local Similarity   73.8%;  Pred. No. 2.7e+02;
  Matches   31;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaagaata 42
          |||| ||| | |||||||| | |   ||||||| ||| || |
Db    382 TATTTTTTCCCCAAAATTTACCTTTTTTTGAGAAAAAAAAA 423


RESULT   8
AQ751027/c
LOCUS       AQ751027      502 bp    DNA             GSS       19-JUL-1999
DEFINITION  HS_5576_B1_B09_SP6 RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1152 Col=17 Row=D, DNA sequence.
ACCESSION   AQ751027
VERSION     AQ751027.1  GI:5538185
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 502)
```

```
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1152  row: D  column: 17
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 502.
FEATURES             Location/Qualifiers
     source          1. .502
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=1152 Col=17 Row=D"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      122 a    118 c     71 g    191 t
ORIGIN


  Query Match             58.1%;  Score 24.4;  DB 211;  Length 502;
  Best Local Similarity   73.8%;  Pred. No. 2.7e+02;
  Matches   31;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaagaata 42
          | || ||| |||||||   | | |||||||| | ||| ||||
Db    423 TTTTTTTTTCTCAAAAAAAGGTAAAATTTTGGGGAAAAAATA 382


RESULT  9
AV718044
LOCUS       AV718044      714 bp    mRNA            EST       16-OCT-2000
DEFINITION  AV718044 FHTA Homo sapiens cDNA clone FHTAABH08 5', mRNA sequence.
ACCESSION   AV718044
VERSION     AV718044.1  GI:10815196
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 714)
  AUTHORS   Xiao,H., Peng,Y., Song,H., Gu,Y., Yang,Y., Gao,G., Xu,X., Li,N.,
            Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu
            ,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Chen,Z. and Han
            ,Z.
  TITLE     Homo sapiens cDNA FHTA clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .714
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="FHTAABH08"
                     /clone_lib="FHTA"
                     /tissue_type="hypothalamus"
                     /dev_stage="Fetal"
                     /lab_host="BM25.8"
                     /note="Vector: pTriplEx2; Site_1: sfiIA; Site_2: sfiIB1"
BASE COUNT      240 a     99 c    121 g    249 t      5 others
ORIGIN


  Query Match             58.1%;  Score 24.4;  DB 32;  Length 714;
  Best Local Similarity   73.8%;  Pred. No. 2.8e+02;
  Matches   31;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaagaata 42
          |||| || | ||||||||| |||||   | || ||||| ||
Db    150 TATTAATTACACAAAATTTGGTCAAAGCATTAGAAAAGATTA 191


RESULT  10
AV718044/c
LOCUS       AV718044      714 bp    mRNA            EST       16-OCT-2000
DEFINITION  AV718044 FHTA Homo sapiens cDNA clone FHTAABH08 5', mRNA sequence.
ACCESSION   AV718044
VERSION     AV718044.1  GI:10815196
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 714)
  AUTHORS   Xiao,H., Peng,Y., Song,H., Gu,Y., Yang,Y., Gao,G., Xu,X., Li,N.,
            Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu,S., Gu
            ,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Chen,Z. and Han
            ,Z.
  TITLE     Homo sapiens cDNA FHTA clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .714
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="FHTAABH08"
                     /clone_lib="FHTA"
                     /tissue_type="hypothalamus"
                     /dev_stage="Fetal"
                     /lab_host="BM25.8"
                     /note="Vector: pTriplEx2; Site_1: sfiIA; Site_2: sfiIB1"
BASE COUNT      240 a     99 c    121 g    249 t      5 others
ORIGIN


  Query Match             58.1%;  Score 24.4;  DB 32;  Length 714;
  Best Local Similarity   73.8%;  Pred. No. 2.8e+02;
  Matches   31;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaagaata 42
          || ||||| || |   ||||| |||||||||| | ||  ||||
Db    191 TAATCTTTTCTAATGCTTTGACCAAATTTTGTGTAATTAATA 150
```

```
RESULT 11
AQ327650
LOCUS       AQ327650      811 bp    DNA             GSS       08-JAN-1999
DEFINITION  nbxb0041B15f CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0041B15f, DNA sequence.
ACCESSION   AQ327650
VERSION     AQ327650.1  GI:4119500
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
            Oryza.
REFERENCE   1  (bases 1 to 811)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence stop: 439.
FEATURES             Location/Qualifiers
     source          1. .811
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="nbxb0041B15f"
                     /clone_lib="CUGI Rice BAC Library"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Rice is one of two most popular grains in the
                     world. Half of the world  population especially those
                     inhabiting highly populated areas of the  humid tropics
                     and subtropics, rely on rice as their primary source of
                     carbohydrate. Monocotyledonous rice is a diploid plant
                     (2n=24) with a  haploid genome equivalent of 431 Mbp
                     (Arumuganathan and Earle, 1991).  The relatively small
                     genome of rice, three times larger than that of
                     Arabidopsis, makes it suitable for genomic studies. In
                     order to  facilitate positional cloning, physical mapping
                     and genome sequencing  of rice, we have constructed a BAC
                     library from Oryza sativa, Nipponbare  variety. The
                     library contains 36,864 clones with an average insert size
                     of 128.5 Kb providing 10.9 haploid genome equivalents. The
                     deep  coverage allows the isolation a particular sequence
                     with a probability  of 99.9 %. Two high density filters,
                     each containing 18,432 clones  (doubly spotted), represent
                     the whole library for colony screening."
BASE COUNT      231 a    144 c    148 g    288 t
ORIGIN


  Query Match             58.1%;  Score 24.4;  DB 205;  Length 811;
  Best Local Similarity   73.8%;  Pred. No. 2.8e+02;
  Matches   31;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaagaata 42
          || | ||| ||| |||||| ||||| |||  || || | |||
Db    512 TAATTTTTTCTCTAAATTTAATCAAGTTTATAGAAATGTATA 553


RESULT 12
AQ327650/c
LOCUS       AQ327650      811 bp    DNA             GSS       08-JAN-1999
DEFINITION  nbxb0041B15f CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0041B15f, DNA sequence.
ACCESSION   AQ327650
VERSION     AQ327650.1  GI:4119500
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
            Oryza.
REFERENCE   1  (bases 1 to 811)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence stop: 439.
FEATURES             Location/Qualifiers
     source          1. .811
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="nbxb0041B15f"
                     /clone_lib="CUGI Rice BAC Library"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Rice is one of two most popular grains in the
                     world. Half of the world  population especially those
                     inhabiting highly populated areas of the  humid tropics
                     and subtropics, rely on rice as their primary source of
                     carbohydrate. Monocotyledonous rice is a diploid plant
                     (2n=24) with a  haploid genome equivalent of 431 Mbp
                     (Arumuganathan and Earle, 1991).  The relatively small
                     genome of rice, three times larger than that of
                     Arabidopsis, makes it suitable for genomic studies. In
                     order to  facilitate positional cloning, physical mapping
                     and genome sequencing  of rice, we have constructed a BAC
                     library from Oryza sativa, Nipponbare  variety. The
                     library contains 36,864 clones with an average insert size
                     of 128.5 Kb providing 10.9 haploid genome equivalents. The
                     deep  coverage allows the isolation a particular sequence
                     with a probability  of 99.9 %. Two high density filters,
                     each containing 18,432 clones  (doubly spotted), represent
                     the whole library for colony screening."
BASE COUNT      231 a    144 c    148 g    288 t
ORIGIN


  Query Match             58.1%;  Score 24.4;  DB 205;  Length 811;
  Best Local Similarity   73.8%;  Pred. No. 2.8e+02;
  Matches   31;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaagaata 42
          ||| | || ||  ||| ||||| |||||| ||| ||| | ||
Db    553 TATACATTTCTATAAACTTGATTAAATTTAGAGAAAAAATTA 512


RESULT 13
H55810
LOCUS       H55810        344 bp    mRNA            EST       02-OCT-1995
DEFINITION  yr02c11.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:204116 5', mRNA sequence.
ACCESSION   H55810
VERSION     H55810.1  GI:1004454
```

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 344)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 293
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: M13RP1
            High quality sequence stop: 293.
FEATURES             Location/Qualifiers
     source          1. .344
                     /organism="Homo sapiens"
                     /db_xref="GDB:3773247"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:204116"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       87 a     51 c     42 g    162 t      2 others
ORIGIN
```

```
  Query Match             57.6%;  Score 24.2;  DB 153;  Length 344;
  Best Local Similarity   78.4%;  Pred. No. 3.1e+02;
  Matches   29;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaa 37
          ||| |||||||||| ||||| || ||| |||  | |||
Db    128 TATCCTTTGCTCATAATTTTATTAAACTTTTTGAAAA 164
```

```
RESULT  14
H55810/c
LOCUS       H55810         344 bp    mRNA            EST       02-OCT-1995
DEFINITION  yr02c11.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:204116 5', mRNA sequence.
ACCESSION   H55810
VERSION     H55810.1  GI:1004454
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 344)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 293
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: M13RP1
            High quality sequence stop: 293.
FEATURES             Location/Qualifiers
     source          1. .344
                     /organism="Homo sapiens"
                     /db_xref="GDB:3773247"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:204116"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       87 a     51 c     42 g    162 t      2 others
ORIGIN
```

```
  Query Match             57.6%;  Score 24.2;  DB 153;  Length 344;
  Best Local Similarity   78.4%;  Pred. No. 3.1e+02;
  Matches   29;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      6 tttgctcaaaatttgatcaaattttgagcaaagaata 42
          ||| |  ||| ||| || ||||| |||||||||| |||
Db    164 TTTTCAAAAAGTTTAATAAAATTATGAGCAAAGGATA 128
```

```
RESULT  15
BF052039
LOCUS       BF052039       456 bp    mRNA            EST       16-OCT-2000
DEFINITION  EST437286 tomato developing/immature green fruit Lycopersicon
            esculentum cDNA clone cLEM25I4 5' sequence similar to ESTs
            gb|T20423, gb|AA712864, gb|H76323 and gb|Z25560 come from this
            gene. {Arabidopsis thaliana}, mRNA sequence.
ACCESSION   BF052039
VERSION     BF052039.1  GI:10805935
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
            I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
            Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
            Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
            ,S.D.
  TITLE     Generation of ESTs from tomato fruit tissue, immature green
  JOURNAL   Unpublished (2000)
COMMENT     Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
```

```
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU.
FEATURES              Location/Qualifiers
     source           1. .456
                      /organism="Lycopersicon esculentum"
                      /cultivar="TA496"
                      /db_xref="taxon:4081"
                      /clone="cLEM25I4"
                      /clone_lib="tomato developing/immature green fruit"
                      /tissue_type="fruit"
                      /dev_stage="immature green (5-35 days post-anthesis)"
                      /lab_host="SOLR"
                      /note="Vector: pBluescriptSKmCUadapt; Site_1: EcoR1;
                      Site_2: Xho1; Fruit were tagged at 5 dpa (0.5 cm) and
                      harvested at 7 day intervals through 35 dpa.  Equal masses
                      of tissue from each stage were combined (including seeds
                      and locules) prior to mRNA isolation."
BASE COUNT       113 a    149 c     46 g    148 t
ORIGIN


  Query Match             57.1%;  Score 24;  DB 143;  Length 456;
  Best Local Similarity   75.0%;  Pred. No. 3.6e+02;
  Matches   30;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1 tattctttgctcaaaatttgatcaaattttgagcaaagaa 40
           | |||| || | ||||   |||| |||||||| | |||||
Db      27 TCTTCTCTGGTGAAAAAAAGATCCAATTTTGAACGAAGAA 66


RESULT  16
BF052039/c
LOCUS       BF052039      456 bp    mRNA            EST       16-OCT-2000
DEFINITION  EST437286 tomato developing/immature green fruit Lycopersicon
            esculentum cDNA clone cLEM25I4 5' sequence similar to ESTs
            gb|T20423, gb|AA712864, gb|H76323 and gb|Z25560 come from this
            gene. {Arabidopsis thaliana}, mRNA sequence.
ACCESSION   BF052039
VERSION     BF052039.1  GI:10805935
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
            I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
            Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
            Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
            ,S.D.
  TITLE     Generation of ESTs from tomato fruit tissue, immature green
  JOURNAL   Unpublished (2000)
COMMENT     Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU.
FEATURES              Location/Qualifiers
     source           1. .456
                      /organism="Lycopersicon esculentum"
                      /cultivar="TA496"
                      /db_xref="taxon:4081"
                      /clone="cLEM25I4"
                      /clone_lib="tomato developing/immature green fruit"
                      /tissue_type="fruit"
                      /dev_stage="immature green (5-35 days post-anthesis)"
                      /lab_host="SOLR"
                      /note="Vector: pBluescriptSKmCUadapt; Site_1: EcoR1;
                      Site_2: Xho1; Fruit were tagged at 5 dpa (0.5 cm) and
                      harvested at 7 day intervals through 35 dpa.  Equal masses
                      of tissue from each stage were combined (including seeds
                      and locules) prior to mRNA isolation."
BASE COUNT       113 a    149 c     46 g    148 t
ORIGIN


  Query Match             57.1%;  Score 24;  DB 143;  Length 456;
  Best Local Similarity   75.0%;  Pred. No. 3.6e+02;
  Matches   30;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       3 ttctttgctcaaaatttgatcaaattttgagcaaagaata 42
           ||||| | |||||||| ||||   |||| | || |||| |
Db      66 TTCTTCGTTCAAAATTGGATCTTTTTTTCACCAGAGAAGA 27


RESULT  17
BG135705
LOCUS       BG135705      509 bp    mRNA            EST       31-JAN-2001
DEFINITION  EST468597 tomato crown gall Lycopersicon esculentum cDNA clone
            cTOE23K4 5' sequence, mRNA sequence.
ACCESSION   BG135705
VERSION     BG135705.1  GI:12635893
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
            I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   van der Hoeven,R., Sun,H., Cho,J., Utterback,T., Hansen,C., Ronning
            ,C. and Tanksley,S.
  TITLE     Generation of ESTs from tomato crown gall tissue
  JOURNAL   Unpublished (2001)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES              Location/Qualifiers
     source           1. .509
                      /organism="Lycopersicon esculentum"
                      /cultivar="TA496"
                      /db_xref="taxon:4081"
                      /clone="cTOE23K4"
                      /clone_lib="tomato crown gall"
                      /tissue_type="crown gall"
                      /dev_stage="crown galls from full-grown plants (8 wks old
                      )"
                      /lab_host="SOLR"
                      /note="Vector: pBluescript SK(-); Site_1: EcoR1; Site_2:
                      Xho1; Four wk old greenhouse plants were stab inoculated
                      on stem with Agrobacterium tumefaciens C58 (Dr. T.J. Burr,
                      Cornell U.). Galls were allowed to develop for another 4
                      wks, when gall tissue was frozen in  liquid nitrogen."
BASE COUNT       119 a    164 c     56 g    170 t
ORIGIN


  Query Match             57.1%;  Score 24;  DB 173;  Length 509;
  Best Local Similarity   75.0%;  Pred. No. 3.7e+02;
  Matches   30;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1 tattctttgctcaaaatttgatcaaattttgagcaaagaa 40
           | |||| || | ||||   |||| |||||||| | |||||
Db      31 TCTTCTCTGGTGAAAAAAAGATCCAATTTTGAACGAAGAA 70


RESULT  18
BG135705/c
LOCUS       BG135705      509 bp    mRNA            EST       31-JAN-2001
DEFINITION  EST468597 tomato crown gall Lycopersicon esculentum cDNA clone
            cTOE23K4 5' sequence, mRNA sequence.
ACCESSION   BG135705
VERSION     BG135705.1  GI:12635893
```

```
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
            I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   van der Hoeven,R., Sun,H., Cho,J., Utterback,T., Hansen,C., Ronning
            ,C. and Tanksley,S.
  TITLE     Generation of ESTs from tomato crown gall tissue
  JOURNAL   Unpublished (2001)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES             Location/Qualifiers
     source          1. .509
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cTOE23K4"
                     /clone_lib="tomato crown gall"
                     /tissue_type="crown gall"
                     /dev_stage="crown galls from full-grown plants (8 wks old
                     )"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; Four wk old greenhouse plants were stab inoculated
                     on stem with Agrobacterium tumefaciens C58 (Dr. T.J. Burr,
                     Cornell U.). Galls were allowed to develop for another 4
                     wks, when gall tissue was frozen in  liquid nitrogen."
BASE COUNT      119 a    164 c     56 g    170 t
ORIGIN


  Query Match             57.1%;  Score 24;  DB 173;  Length 509;
  Best Local Similarity   75.0%;  Pred. No. 3.7e+02;
  Matches   30;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      3 ttctttgctcaaaatttgatcaaattttgagcaaagaata 42
          ||||| | |||||||| ||||   |||| | || |||| |
Db     70 TTCTTCGTTCAAAATTGGATCTTTTTTTCACCAGAGAAGA 31


RESULT  19
AW928999
LOCUS       AW928999      557 bp    mRNA            EST       30-MAY-2000
DEFINITION  EST337883 tomato flower buds 8 mm to pre-anthesis, Cornell
            University Lycopersicon esculentum cDNA clone cTOC4O4 5', mRNA
            sequence.
ACCESSION   AW928999
VERSION     AW928999.1  GI:8104496
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
            I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 557)
  AUTHORS   van der Hoeven,R.S., Bezzerides,J.l., Matern,A.L., Holt,I.E., Liang
            ,F., Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning
            ,C.M., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
  TITLE     Generation of ESTs from tomato flower tissue
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU
            5 prime sequence.
FEATURES             Location/Qualifiers
     source          1. .557
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cTOC4O4"
                     /clone_lib="tomato flower buds 8 mm to pre-anthesis,
                     Cornell University"
                     /tissue_type="flower"
                     /dev_stage="buds 8mm-to-preanthesis"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; supplier: Tanksley;  Flower buds and flowers were
                     taken from greenhouse plants (4-8 wks old, TA496). They
                     were immediately frozen in liquid nitrogen and then
                     size-separated while remaining frozen."
BASE COUNT      135 a    173 c     64 g    185 t
ORIGIN


  Query Match             57.1%;  Score 24;  DB 121;  Length 557;
  Best Local Similarity   75.0%;  Pred. No. 3.7e+02;
  Matches   30;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 tattctttgctcaaaatttgatcaaattttgagcaaagaa 40
          | |||| || | ||||   |||| |||||||| | |||||
Db     31 TCTTCTCTGGTGAAAAAAAGATCCAATTTTGAACGAAGAA 70


RESULT  20
AW928999/c
LOCUS       AW928999      557 bp    mRNA            EST       30-MAY-2000
DEFINITION  EST337883 tomato flower buds 8 mm to pre-anthesis, Cornell
            University Lycopersicon esculentum cDNA clone cTOC4O4 5', mRNA
            sequence.
ACCESSION   AW928999
VERSION     AW928999.1  GI:8104496
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
            I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 557)
  AUTHORS   van der Hoeven,R.S., Bezzerides,J.l., Matern,A.L., Holt,I.E., Liang
            ,F., Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning
            ,C.M., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
  TITLE     Generation of ESTs from tomato flower tissue
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU
            5 prime sequence.
FEATURES             Location/Qualifiers
     source          1. .557
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cTOC4O4"
                     /clone_lib="tomato flower buds 8 mm to pre-anthesis,
                     Cornell University"
                     /tissue_type="flower"
                     /dev_stage="buds 8mm-to-preanthesis"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; supplier: Tanksley;  Flower buds and flowers were
                     taken from greenhouse plants (4-8 wks old, TA496). They
                     were immediately frozen in liquid nitrogen and then
                     size-separated while remaining frozen."
BASE COUNT      135 a    173 c     64 g    185 t
ORIGIN
```

Query Match . 57.1%; Score 24; DB 121; Length 557;
Best Local Similarity 75.0%; Pred. No. 3.7e+02;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy       3 ttctttgctcaaaatttgatcaaattttgagcaaagaata 42
           ||||| | |||||||| ||||    |||| | || |||| |
Db      70 TTCTTCGTTCAAAATTGGATCTTTTTTTCACCAGAGAAGA 31
```

Search completed: June 7, 2001, 17:41:30
Job time: 5368 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 7, 2001, 18:05:31 ; Search time 1335.93 Seconds
(without alignments)
419.497 Million cell updates/sec

Title: US-09-601-561-3
Perfect score: 38
Sequence: 1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		23.8	62.6	45770	5	CELR05C11	AF125446 Caenorhab
2	c	23.8	62.6	45770	5	CELR05C11	AF125446 Caenorhab
3		23.8	62.6	188972	60	AC006907	AC006907 Caenorhab
4	c	23.8	62.6	188972	60	AC006907	AC006907 Caenorhab
5		23.6	62.1	2637	2	CHRFBFPGE	X91081 C.hyoilei r
6	c	23.6	62.1	2637	2	CHRFBFPGE	X91081 C.hyoilei r
7		23.6	62.1	159974	89	AL139008	AL139008 Human DNA
8	c	23.6	62.1	159974	89	AL139008	AL139008 Human DNA
9		23.6	62.1	169223	79	AL353675	AL353675 Homo sapi
10	c	23.6	62.1	169223	79	AL353675	AL353675 Homo sapi
11		23.2	61.1	164	56	SYNFLUAPB	M25075 Plasmid vec

c	12	23.2	61.1	164	56	SYNFLUAPB	M25075 Plasmid vec
	13	22.8	60.0	144402	67	AC021854	AC021854 Homo sapi
c	14	22.8	60.0	144402	67	AC021854	AC021854 Homo sapi
	15	22.8	60.0	171549	69	AC024662	AC024662 Homo sapi
c	16	22.8	60.0	171549	69	AC024662	AC024662 Homo sapi
	17	22.8	60.0	196721	87	AC019179	AC019179 Homo sapi
c	18	22.8	60.0	196721	87	AC019179	AC019179 Homo sapi
	19	22.6	59.5	4267	7	SSU70240	U70240 Sus scrofa
c	20	22.6	59.5	4267	7	SSU70240	U70240 Sus scrofa
	21	22.6	59.5	167547	73	AC067725	AC067725 Homo sapi
c	22	22.6	59.5	167547	73	AC067725	AC067725 Homo sapi
	23	22.6	59.5	170616	77	AC084408	AC084408 Homo sapi
c	24	22.6	59.5	170616	77	AC084408	AC084408 Homo sapi
	25	22.4	58.9	92611	86	AC006398	AC006398 Homo sapi
c	26	22.4	58.9	92611	86	AC006398	AC006398 Homo sapi
	27	22.4	58.9	162591	90	AL391839	AL391839 Human DNA
c	28	22.4	58.9	162591	90	AL391839	AL391839 Human DNA
	29	22.4	58.9	172148	75	AC073132	AC073132 Homo sapi
c	30	22.4	58.9	172148	75	AC073132	AC073132 Homo sapi
	31	22.2	58.4	205141	65	AC018679	AC018679 Homo sapi
c	32	22.2	58.4	205141	65	AC018679	AC018679 Homo sapi
	33	22	57.9	40457	5	CELF54D7	AF039712 Caenorhab
c	34	22	57.9	40457	5	CELF54D7	AF039712 Caenorhab
	35	22	57.9	139407	80	AL356127	AL356127 Homo sapi
c	36	22	57.9	139407	80	AL356127	AL356127 Homo sapi
	37	22	57.9	159075	72	AC036224	AC036224 Homo sapi
c	38	22	57.9	159075	72	AC036224	AC036224 Homo sapi
	39	22	57.9	163147	69	AC024445	AC024445 Homo sapi
c	40	22	57.9	163147	69	AC024445	AC024445 Homo sapi
	41	22	57.9	169075	62	AC011940	AC011940 Homo sapi
c	42	22	57.9	169075	62	AC011940	AC011940 Homo sapi
	43	22	57.9	175184	87	AC022367	AC022367 Homo sapi
c	44	22	57.9	175184	87	AC022367	AC022367 Homo sapi
	45	22	57.9	179975	69	AC025185	AC025185 Homo sapi
c	46	22	57.9	179975	69	AC025185	AC025185 Homo sapi
	47	22	57.9	200889	69	AC024191	AC024191 Homo sapi
	48	22	57.9	210730	64	AC015975	AC015975 Homo sapi
c	49	22	57.9	210730	64	AC015975	AC015975 Homo sapi
	50	22	57.9	257728	60	AC006846	AC006846 Caenorhab

ALIGNMENTS

RESULT 1
CELR05C11
LOCUS CELR05C11 45770 bp DNA INV 06-FEB-1999
DEFINITION Caenorhabditis elegans cosmid R05C11.
ACCESSION AF125446
VERSION AF125446.1 GI:4226080
KEYWORDS .
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 45770)
AUTHORS Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
Wilkinson-Sproat,J. and Wohldman,P.
TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 45770)
AUTHORS Lamar,B., Kramer,J. and Gibson,A.
TITLE The sequence of C. elegans cosmid R05C11
JOURNAL Unpublished (1999)
REFERENCE 3 (bases 1 to 45770)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 IRQ, England
e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' YAC segment is Y76B12C; 3' YAC segment is Y76B12B, 200 bp
overlap. Actual start of this clone is at 1 of CELR05C11; actual
end is at 43876 of CELR05C11.

NOTES:

Coding sequences below are predicted from computer analysis, using
the program Genefinder(P. Green and L. Hillier, ms in preparation).

FEATURES Location/Qualifiers
source 1. .45770
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/clone="R05C11"
/chromosome="IV"
gene complement(11707. .12940)
/gene="R05C11.2"
CDS complement(join(11707. .11985,12032. .12776,12825. .12940))
/gene="R05C11.2"
/codon_start=1
/evidence=not_experimental
/protein_id="AAD12805.1"
/db_xref="GI:4226082"
/translation="MDETKLLSEVSGSSISSTLWFYNTVIMNSVVERLDFFDIQVLRK
ISYGSRQCIDLIKPDHKLLKIGIHVQSIGCALLEMFTESTNRKVSYKASRLSRKSSEL
DYGNFSNDLFLILKNQKSALEELKMNFMQVDETSESSFKDVTEEEKKYMDEIEGNPQE
TESHSIWDQNENRFLKSFKDLHASLGRLLPVEKLVVTTRNQKTIVQLLQSVDPSKLNK
IEFGTSFDVPASDVWSIEEFLVMEQWRNAKELTIIKRVIKCDVENLVHFRKVDIILDS
LSLKNVIYLRDMFLVSPDFLQFKVTFVNSDVNNSIFQEGFLGIPFRIVTDSTNNQPSI
WFFNCMNDTDDLQIRYYQNLKKFMFNRIPRESSPAIPRESSPARC"
gene 22596. .25485
/gene="R05C11.1"
CDS join(22596. .22650,22702. .22784,24373. .24519,24572. .24688,
25276. .25485)
/gene="R05C11.1"
/codon_start=1
/evidence=not_experimental
/protein_id="AAD12804.1"
/db_xref="GI:4226081"
/translation="MCTALMAIPEDGDTEVVSEKTFKVFSLLNLDAPYLHDKAHVTEE
EEITGCCGRRKGALIATMQVDEIGVMPGMTSKISLTVENKTKKRKKWMRKKDNHECVL
LSLCQQLDFVAVNRNDPMMIDRKSITIAVESHGTCKTKAGVGPQTKEIEFSVPPNLPP

```
                TSIHANRLVTVSYFFKLDLDHFDLVLPVIVGSVKSSASNGSEK"
     misc_feature    37862. .37863
                     /note="SL1 trans-splice site; see GB:AJ010708"
     gene            37880. .45233
                     /gene="R05C11.3"
     CDS             join(37880. .37963,38019. .38277,38666. .38785,38830. .39440,
                     39485. .39737,40376. .41374,41990. .42552,43293. .43457,
                     43565. .43771,45162. .45233)
                     /gene="R05C11.3"
                     /note="C. elegans calcium ATPase MCA-2 (GB:AJ010708);
                     similar to E1-E2 ATPases (Pfam:PF00122), score=384.8,
                     E=8.9e-112, N=1); coded for by C. elegans cDNA yk68b10.5;
                     coded for by C. elegans cDNA yk44a6.5; coded for by C.
                     elegans cDNA yk146a1.5; coded for by C. elegans cDNA
                     yk396d12.5; coded for by C. elegans cDNA yk225b3.5; coded
                     for by C. elegans cDNA yk44a6.3; coded for by C. elegans
                     cDNA yk146a1.3; coded for by C. elegans cDNA yk225b3.3;
                     coded for by C. elegans cDNA yk396d12.3"
                     /codon_start=1
                     /protein_id="AAD12806.1"
                     /db_xref="GI:4226083"
                     /translation="MPSSDDYDCSISELKLLMELRGAEALEKVNSTYNGVEGLCRKLK
                     TDPINGLPNDTKELEHRRTAFGKNEIPPAPSKSFFRLAWEALQDITLVILLVAALVSL
                     GLSFYKPPAEHASNDSSESEAGWIEGVAILVAVLVVVLVTALNDWTKEKQFRGLQSKI
                     ETEHKFSVIRGGEPLDIVVNELVVGDIARVKYGDLLPADGLLIQSNDLKIDESSLTGE
                     SDLIRKSEEFDPVLLSGTHAMEGSGRFLVTAVGLNSQTGIIMSLLGAAKEKKDDKKEE
                     PTTLTNGNHMNGLGNGVDKAAELPVPEEEEVGRMSKSVLQTKLSNLALQIGYIGSIVA
                     AATVLILIIRHCISKYAIEGKSFEASDISHFVNFIIIGVTVLVIAVPEGLPLAITLAL
                     TYSVKKMMKDNNLVRHLDACETMGNATSICSDKTGTLTTNRMTCVQQYINSEFYKGNA
                     PKYEQMDPSTRDILFNGIVINSGYNSTVVTPKNPGEQRGQIGNKTECSLLGFIMDSGR
                     SYEDLRRQFPEEKLYKVYTFNSSRKSMMTVIELGDKKYRIYAKGASEIILTRCNYIFG
                     KSGKIEQFGPKEAAVMTKNVIEPMASDGLRTIGLAFKDLVPAGSKKYEYEEEYDGEID
                     WEDEEKIREGQTVIAVMGIQDPVRPEVPAAIAKCQKAGITVRMVTGDNINTARSIATQ
                     CGIMQPGGDFLALEGKEFNARIRDADGKVNQQKFDAIWPKLRVLARAQPSDKYVLVKG
                     IIESTVTKNREVVAVTGDGTNDAPALKKADVGFAMGIAGTDVAKEASDIILTDDNFSS
                     IVKAVMWGRNVYDSIAKFLQFQLTVNVVAVTIAFIGACAISDSPLKAVQMLWVNLIMD
                     TLASLALATEMPTEDLLNRKPYGRTKSLISRTMVKNIVGHAVYQLAILFAIMFWGDKL
                     IPNTPSGRNAELGSPPSAHFTIIFNAFVLMTLVNEINARKIHGERNVFKGIFTNPIFC
                     VIWITTLISHILIVQFGGQWFSTAPLDLTQWIICICCGVGELFWGQIINCIPASILPK
                     SFRFGKGDVQPTSIMLSGEYDMPSTASTLPMKDGAAADHQRAGQSLWLLGLTRLQTQG
                     LVAYEPVDRYILRRQMQRIHPE"
BASE COUNT    14313 a   8643 c   7686 g  15128 t
ORIGIN


  Query Match             62.6%;  Score 23.8;  DB 5;  Length 45770;
  Best Local Similarity   80.0%;  Pred. No. 1.7e+02;
  Matches   28;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3 gcttttgctcaaagtttaaactttgagcaaaagcc 37
           | |||||||| || ||| ||  |||||||||| ||
Db    7334 GGTTTTGCTCGAATTTTCAAAATTGAGCAAAACCC 7368


RESULT   2
CELR05C11/c
LOCUS       CELR05C11   45770 bp    DNA             INV       06-FEB-1999
DEFINITION  Caenorhabditis elegans cosmid R05C11.
ACCESSION   AF125446
VERSION     AF125446.1  GI:4226080
KEYWORDS    .
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 45770)
  AUTHORS   Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
            Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
            Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
            Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
            Wilkinson-Sproat,J. and Wohldman,P.
    TITLE   2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 45770)
  AUTHORS   Lamar,B., Kramer,J. and Gibson,A.
    TITLE   The sequence of C. elegans cosmid R05C11
  JOURNAL   Unpublished (1999)
REFERENCE   3  (bases 1 to 45770)
  AUTHORS   Waterston,R.
    TITLE   Direct Submission
  JOURNAL   Submitted (03-FEB-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     Submitted by:
                Genome Sequencing Center
                Department of Genetics, Washington University,
                St. Louis, MO 63110, USA, and
                Sanger Centre, Hinxton Hall
                Cambridge CB10 IRQ, England
                e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk


            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone

                        NEIGHBORING COSMID INFORMATION:


            The 5' YAC segment is Y76B12C; 3' YAC segment is Y76B12B, 200 bp
            overlap.  Actual start of this clone is at 1 of CELR05C11; actual
            end is at 43876 of CELR05C11.

            NOTES:

            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES                Location/Qualifiers
     source             1. .45770
                        /organism="Caenorhabditis elegans"
                        /strain="Bristol N2"
                        /db_xref="taxon:6239"
                        /clone="R05C11"
                        /chromosome="IV"
     gene               complement(11707. .12940)
                        /gene="R05C11.2"
     CDS                complement(join(11707. .11985,12032. .12776,12825. .12940))
                        /gene="R05C11.2"
                        /codon_start=1
                        /evidence=not_experimental
                        /protein_id="AAD12805.1"
                        /db_xref="GI:4226082"
                        /translation="MDETKLLSEVSGSSISSTLWFYNTVIMNSVVERLDFFDIQVLRK
                        ISYGSRQCIDLIKPDHKLLKIGIHVQSIGCALLEMFTESTNRKVSYKASRLSRKSSEL
                        DYGNFSNDLFLILKNQKSALEELKMNFMQVDETSESSFKDVTEEEKKYMDEIEGNPQE
                        TESHSIWDQNENRFLKSFKDLHASLGRLLPVEKLVVTTRNQKTIVQLLQSVDPSKLNK
                        IEFGTSFDVPASDVWSIEEFLVMEQWRNAKELTIIKRVIKCDVENLVHFRKVDIILDS
                        LSLKNVIYLRDMFLVSPDFLQFKVTFVNSDVNNSIFQEGFLGIPFRIVTDSTNNQPSI
                        WFFNCMNDTDDLQIRYYQNLKKFMFNRIPRESSPAIPRESSPARC"
     gene               22596. .25485
                        /gene="R05C11.1"
```

```
     CDS             join(22596. .22650,22702. .22784,24373. .24519,24572. .24688,
                     25276. .25485)
                     /gene="R05C11.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAD12804.1"
                     /db_xref="GI:4226081"
                     /translation="MCTALMAIPEDGDTEVVSEKTFKVFSLLNLDAPYLHDKAHVTEE
                     EEITGCCGRRKGALIATMQVDEIGVMPGMTSKISLTVENKTKKRKKWMRKKDNHECVL
                     LSLCQQLDFVAVNRNDPMMIDRKSITIAVESHGTCKTKAGVGPQTKEIEFSVPPNLPP
                     TSIHANRLVTVSYFFKLDLDHFDLVLPVIVGSVKSSASNGSEK"
     misc_feature    37862. .37863
                     /note="SL1 trans-splice site; see GB:AJ010708"
     gene            37880. .45233
                     /gene="R05C11.3"
     CDS             join(37880. .37963,38019. .38277,38666. .38785,38830. .39440,
                     39485. .39737,40376. .41374,41990. .42552,43293. .43457,
                     43565. .43771,45162. .45233)
                     /gene="R05C11.3"
                     /note="C. elegans calcium ATPase MCA-2 (GB:AJ010708);
                     similar to E1-E2 ATPases (Pfam:PF00122), score=384.8,
                     E=8.9e-112, N=1); coded for by C. elegans cDNA yk68b10.5;
                     coded for by C. elegans cDNA yk44a6.5; coded for by C.
                     elegans cDNA yk146a1.5; coded for by C. elegans cDNA
                     yk396d12.5; coded for by C. elegans cDNA yk225b3.5; coded
                     for by C. elegans cDNA yk44a6.3; coded for by C. elegans
                     cDNA yk146a1.3; coded for by C. elegans cDNA yk225b3.3;
                     coded for by C. elegans cDNA yk396d12.3"
                     /codon_start=1
                     /protein_id="AAD12806.1"
                     /db_xref="GI:4226083"
                     /translation="MPSSDDYDCSISELKLLMELRGAEALEKVNSTYNGVEGLCRKLK
                     TDPINGLPNDTKELEHRRTAFGKNEIPPAPSKSFFRLAWEALQDITLVILLVAALVSL
                     GLSFYKPPAEHASNDSSESEAGWIEGVAILVAVLVVVLVTALNDWTKEKQFRGLQSKI
                     ETEHKFSVIRGGEPLDIVVNELVVGDIARVKYGDLLPADGLLIQSNDLKIDESSLTGE
                     SDLIRKSEEFDPVLLSGTHAMEGSGRFLVTAVGLNSQTGIIMSLLGAAKEKKDDKKEE
                     PTTLTNGNHMNGLGNGVDKAAELPVPEEEEVGRMSKSVLQTKLSNLALQIGYIGSIVA
                     AATVLILIIRHCISKYAIEGKSFEASDISHFVNFIIIGVTVLVIAVPEGLPLAITLAL
                     TYSVKKMMKDNNLVRHLDACETMGNATSICSDKTGTLTTNRMTCVQQYINSEFYKGNA
                     PKYEQMDPSTRDILFNGIVINSGYNSTVVTPKNPGEQRGQIGNKTECSLLGFIMDSGR
                     SYEDLRRQFPEEKLYKVYTFNSSRKSMMTVIELGDKKYRIYAKGASEIILTRCNYIFG
                     KSGKIEQFGPKEAAVMTKNVIEPMASDGLRTIGLAFKDLVPAGSKKYEYEEEYDGEID
                     WEDEEKIREGQTVIAVMGIQDPVRPEVPAAIAKCQKAGITVRMVTGDNINTARSIATQ
                     CGIMQPGGDFLALEGKEFNARIRDADGKVNQQKFDAIWPKLRVLARAQPSDKYVLVKG
                     IIESTVTKNREVVAVTGDGTNDAPALKKADVGFAMGIAGTDVAKEASDIILTDDNFSS
                     IVKAVMWGRNVYDSIAKFLQFQLTVNVVAVTIAFIGACAISDSPLKAVQMLWVNLIMD
                     TLASLALATEMPTEDLLNRKPYGRTKSLISRTMVKNIVGHAVYQLAILFAIMFWGDKL
                     IPNTPSGRNAELGSPPSAHFTIIFNAFVLMTLVNEINARKIHGERNVFKGIFTNPIFC
                     VIWITTLISHILIVQFGGQWFSTAPLDLTQWIICICCGVGELFWGQIINCIPASILPK
                     SFRFGKGDVQPTSIMLSGEYDMPSTASTLPMKDGAAADHQRAGQSLWLLGLTRLQTQG
                     LVAYEPVDRYILRRQMQRIHPE"
BASE COUNT    14313 a    8643 c    7686 g   15128 t
ORIGIN


  Query Match             62.6%;  Score 23.8;  DB 5;  Length 45770;
  Best Local Similarity   80.0%;  Pred. No. 1.7e+02;
  Matches   28;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      2 ggcttttgctcaaagtttaaactttgagcaaaagc 36
          || |||||||||  || ||| || ||||||| |
Db   7368 GGGTTTTGCTCAATTTTGAAAATTCGAGCAAAACC 7334


RESULT   3
AC006907
LOCUS       AC006907   188972 bp    DNA             HTG       26-FEB-1999
DEFINITION  Caenorhabditis elegans clone Y76B12, *** SEQUENCING IN PROGRESS
            ***, 25 unordered pieces.
ACCESSION   AC006907
VERSION     AC006907.2  GI:4309897
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 188972)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Caehorhabditis elegans clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 188972)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 1, 1999 this sequence version replaced gi:4263441.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4051: contig of 4051 bp in length
            *     4052     4061: gap of unknown length
            *     4062    15145: contig of 11084 bp in length
            *    15146    15155: gap of unknown length
            *    15156    18371: contig of 3216 bp in length
            *    18372    18381: gap of unknown length
            *    18382    29860: contig of 11479 bp in length
            *    29861    29870: gap of unknown length
            *    29871    40166: contig of 10296 bp in length
            *    40167    40176: gap of unknown length
            *    40177    44947: contig of 4771 bp in length
            *    44948    44957: gap of unknown length
            *    44958    50599: contig of 5642 bp in length
            *    50600    50609: gap of unknown length
            *    50610    59733: contig of 9124 bp in length
            *    59734    59743: gap of unknown length
            *    59744    68416: contig of 8673 bp in length
            *    68417    68426: gap of unknown length
            *    68427    70667: contig of 2241 bp in length
            *    70668    70677: gap of unknown length
            *    70678    87214: contig of 16537 bp in length
            *    87215    87224: gap of unknown length
            *    87225    91568: contig of 4344 bp in length
            *    91569    91578: gap of unknown length
            *    91579    95948: contig of 4370 bp in length
            *    95949    95958: gap of unknown length
            *    95959    98670: contig of 2712 bp in length
            *    98671    98680: gap of unknown length
            *    98681   103766: contig of 5086 bp in length
            *   103767   103776: gap of unknown length
            *   103777   110594: contig of 6818 bp in length
            *   110595   110604: gap of unknown length
            *   110605   117587: contig of 6983 bp in length
            *   117588   117597: gap of unknown length
            *   117598   136489: contig of 18892 bp in length
            *   136490   136499: gap of unknown length
            *   136500   141197: contig of 4698 bp in length
            *   141198   141206: gap of unknown length
            *   141207   147501: contig of 6295 bp in length
            *   147502   147510: gap of unknown length
            *   147511   161272: contig of 13762 bp in length
            *   161273   161281: gap of unknown length
            *   161282   163240: contig of 1959 bp in length
            *   163241   163249: gap of unknown length
            *   163250   167605: contig of 4356 bp in length
            *   167606   167614: gap of unknown length
            *   167615   170377: contig of 2763 bp in length
            *   170378   170386: gap of unknown length
            *   170387   188972: contig of 18586 bp in length.
FEATURES             Location/Qualifiers
     source          1. .188972
```

```
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /clone="Y76B12"
BASE COUNT     61066 a  32767 c  33053 g  60509 t   1577 others
ORIGIN


  Query Match             62.6%;  Score 23.8;  DB 60;  Length 188972;
  Best Local Similarity   80.0%;  Pred. No. 1.3e+02;
  Matches   28;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      3 gcttttgctcaaagtttaaactttgagcaaaagcc 37
          | ||||||||| || ||| ||  |||||||||| ||
Db 100902 GGTTTTGCTCGAATTTTCAAAATTGAGCAAAACCC 100936


RESULT   4
AC006907/c
LOCUS       AC006907   188972 bp    DNA             HTG       26-FEB-1999
DEFINITION  Caenorhabditis elegans clone Y76B12, *** SEQUENCING IN PROGRESS
            ***, 25 unordered pieces.
ACCESSION   AC006907
VERSION     AC006907.2  GI:4309897
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 188972)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Caehorhabditis elegans clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 188972)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 1, 1999 this sequence version replaced gi:4263441.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4051: contig of 4051 bp in length
            *     4052     4061: gap of unknown length
            *     4062    15145: contig of 11084 bp in length
            *    15146    15155: gap of unknown length
            *    15156    18371: contig of 3216 bp in length
            *    18372    18381: gap of unknown length
            *    18382    29860: contig of 11479 bp in length
            *    29861    29870: gap of unknown length
            *    29871    40166: contig of 10296 bp in length
            *    40167    40176: gap of unknown length
            *    40177    44947: contig of 4771 bp in length
            *    44948    44957: gap of unknown length
            *    44958    50599: contig of 5642 bp in length
            *    50600    50609: gap of unknown length
            *    50610    59733: contig of 9124 bp in length
            *    59734    59743: gap of unknown length
            *    59744    68416: contig of 8673 bp in length
            *    68417    68426: gap of unknown length
            *    68427    70667: contig of 2241 bp in length
            *    70668    70677: gap of unknown length
            *    70678    87214: contig of 16537 bp in length
            *    87215    87224: gap of unknown length
            *    87225    91568: contig of 4344 bp in length
            *    91569    91578: gap of unknown length
            *    91579    95948: contig of 4370 bp in length
            *    95949    95958: gap of unknown length
            *    95959    98670: contig of 2712 bp in length
            *    98671    98680: gap of unknown length
            *    98681   103766: contig of 5086 bp in length
            *   103767   103776: gap of unknown length
            *   103777   110594: contig of 6818 bp in length
            *   110595   110604: gap of unknown length
            *   110605   117587: contig of 6983 bp in length
            *   117588   117597: gap of unknown length
            *   117598   136489: contig of 18892 bp in length
            *   136490   136499: gap of unknown length
            *   136500   141197: contig of 4698 bp in length
            *   141198   141206: gap of unknown length
            *   141207   147501: contig of 6295 bp in length
            *   147502   147510: gap of unknown length
            *   147511   161272: contig of 13762 bp in length
            *   161273   161281: gap of unknown length
            *   161282   163240: contig of 1959 bp in length
            *   163241   163249: gap of unknown length
            *   163250   167605: contig of 4356 bp in length
            *   167606   167614: gap of unknown length
            *   167615   170377: contig of 2763 bp in length
            *   170378   170386: gap of unknown length
            *   170387   188972: contig of 18586 bp in length.
FEATURES             Location/Qualifiers
     source          1. .188972
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /clone="Y76B12"
BASE COUNT     61066 a  32767 c  33053 g  60509 t   1577 others
ORIGIN


  Query Match             62.6%;  Score 23.8;  DB 60;  Length 188972;
  Best Local Similarity   80.0%;  Pred. No. 1.3e+02;
  Matches   28;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      2 ggcttttgctcaaagtttaaactttgagcaaaagc 36
            || ||||||||||  || ||| || |||||||| |
Db 100936 GGGTTTTGCTCAATTTTGAAAATTCGAGCAAAACC 100902


RESULT   5
CHRFBFPGE
LOCUS       CHRFBFPGE    2637 bp    DNA             BCT       11-DEC-1997
DEFINITION  C.hyoilei rfbF and rfbP genes.
ACCESSION   X91081
VERSION     X91081.1  GI:1486282
KEYWORDS    galactosyltransferase; rfbF gene; rfbP gene.
SOURCE      Campylobacter hyoilei.
  ORGANISM  Campylobacter hyoilei
            Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
            Campylobacter.
REFERENCE   1  (bases 1 to 2637)
  AUTHORS   Korolik,V., Fry,B.N., Alderton,M.R., van der Zeijst,B.A. and
            Coloe,P.J.
  TITLE     Expression of Campylobacter hyoilei lipo-oligosaccharide (LOS)
            antigens in Escherichia coli
  JOURNAL   Microbiology 143 (Pt 11), 3481-3489 (1997)
  MEDLINE   98048472
REFERENCE   2  (bases 1 to 2637)
  AUTHORS   Fry,B.N.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-1995) B.N. Fry, Infectious Diseases and
            Immunology, Bacteriology, School of Veterinary Medicine, University
            of Utrecht, P.O.Box 80.165, 3508 TD Utrecht, NETHERLANDS
  REMARK    Revised by [3]
REFERENCE   3  (bases 1 to 2637)
  AUTHORS   Fry,B.N.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-1996) B.N. Fry, Infectious Diseases and
            Immunology, Bacteriology, School of Veterinary Medicine, University
            of Utrecht, P.O.Box 80.165, 3508 TD Utrecht, NETHERLANDS
COMMENT     On Aug 9, 1996 this sequence version replaced gi:1050452.
```

```
FEATURES             Location/Qualifiers
     source          1. .2637
                     /organism="Campylobacter hyoilei"
                     /strain="RMIT-32A"
                     /specific_host="pig"
                     /db_xref="taxon:28078"
                     /clone="Lpsch6"
                     /clone="Lpsch11"
     gene            773. .1903
                     /gene="rfbF"
     CDS             773. .1903
                     /gene="rfbF"
                     /codon_start=1
                     /transl_table=11
                     /product="galactosyltransferase"
                     /protein_id="CAA62555.1"
                     /db_xref="GI:1486283"
                     /db_xref="SPTREMBL:P71123"
                     /translation="MRVGFLTHAGASIYHFRLPIIKALIARGDEVFVIVPQDEYTEKL
                     KALNLNIVVYELSRASLNPLIVFKNFLHLKNVLKNLNLDLLQSGAHKSNTFGLIAAKY
                     AKIPYKIGLVEGLGSFYIDKGFKANLVRFVINTLYKLSFKIADSFIFVNQANADFMRN
                     LGLKENKICVIKSVGINLKKFFPMRVEQEAKKAFWQNLKIDEKPIVLMIARALWHKGV
                     KEFYESAEYLKDRANFVLVGGRDDNPSCASLEFLNSGKVFYLGARSDIVELLQNCDIF
                     VLPSYKEGFPVSVLEAKACGKVIVVSDCEGCVEAISNAYDGLWAKTKDSKDLIEKIQV
                     LLEDESLRINLGKNAAKDALQYDENVIAQRYLELYDRVIKNV"
     gene            1896. .2498
                     /gene="rfbP"
     CDS             1896. .2498
                     /gene="rfbP"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA62556.1"
                     /db_xref="GI:1050454"
                     /db_xref="SPTREMBL:P71124"
                     /translation="MYEKWIKRIFDFVLALFLLVLFSPLILITALLLKITQGSVIFTQ
                     NRPGLNEKIFKIYKFKTMSDERDEKGELLSDELRLKAFGKIVRSLSLDELLQLFNVLK
                     GDMSFVGPRPLLVEYLPLYNEEQKLRHKVRPGITGWAQVNGRNAISWQKKFELDVYYV
                     KNISFLLDLKIMFLTALKVLKRSGVSKEGHVTTEKFNGKN"
BASE COUNT      886 a    307 c    516 g    928 t
ORIGIN
```

```
  Query Match             62.1%;  Score 23.6;  DB 2;  Length 2637;
  Best Local Similarity   76.3%;  Pred. No. 3.6e+02;
  Matches   29;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38
          | ||||||   |||| ||||||  ||||  ||||||||
Db   1347 AAGCTTTTTGGCAAAATTTAAAAATTGATGAAAAGCCT 1384
```

```
RESULT   6
CHRFBFPGE/c
LOCUS       CHRFBFPGE    2637 bp    DNA             BCT       11-DEC-1997
DEFINITION  C.hyoilei rfbF and rfbP genes.
ACCESSION   X91081
VERSION     X91081.1  GI:1486282
KEYWORDS    galactosyltransferase; rfbF gene; rfbP gene.
SOURCE      Campylobacter hyoilei.
  ORGANISM  Campylobacter hyoilei
            Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
            Campylobacter.
REFERENCE   1  (bases 1 to 2637)
  AUTHORS   Korolik,V., Fry,B.N., Alderton,M.R., van der Zeijst,B.A. and
            Coloe,P.J.
  TITLE     Expression of Campylobacter hyoilei lipo-oligosaccharide (LOS)
            antigens in Escherichia coli
  JOURNAL   Microbiology 143 (Pt 11), 3481-3489 (1997)
  MEDLINE   98048472
REFERENCE   2  (bases 1 to 2637)
  AUTHORS   Fry,B.N.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-1995) B.N. Fry, Infectious Diseases and
            Immunology, Bacteriology, School of Veterinary Medicine, University
            of Utrecht, P.O.Box 80.165, 3508 TD Utrecht, NETHERLANDS
  REMARK    Revised by [3]
REFERENCE   3  (bases 1 to 2637)
  AUTHORS   Fry,B.N.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-1996) B.N. Fry, Infectious Diseases and
            Immunology, Bacteriology, School of Veterinary Medicine, University
            of Utrecht, P.O.Box 80.165, 3508 TD Utrecht, NETHERLANDS
COMMENT     On Aug 9, 1996 this sequence version replaced gi:1050452.
FEATURES             Location/Qualifiers
     source          1. .2637
                     /organism="Campylobacter hyoilei"
                     /strain="RMIT-32A"
                     /specific_host="pig"
                     /db_xref="taxon:28078"
                     /clone="Lpsch6"
                     /clone="Lpsch11"
     gene            773. .1903
                     /gene="rfbF"
     CDS             773. .1903
                     /gene="rfbF"
                     /codon_start=1
                     /transl_table=11
                     /product="galactosyltransferase"
                     /protein_id="CAA62555.1"
                     /db_xref="GI:1486283"
                     /db_xref="SPTREMBL:P71123"
                     /translation="MRVGFLTHAGASIYHFRLPIIKALIARGDEVFVIVPQDEYTEKL
                     KALNLNIVVYELSRASLNPLIVFKNFLHLKNVLKNLNLDLLQSGAHKSNTFGLIAAKY
                     AKIPYKIGLVEGLGSFYIDKGFKANLVRFVINTLYKLSFKIADSFIFVNQANADFMRN
                     LGLKENKICVIKSVGINLKKFFPMRVEQEAKKAFWQNLKIDEKPIVLMIARALWHKGV
                     KEFYESAEYLKDRANFVLVGGRDDNPSCASLEFLNSGKVFYLGARSDIVELLQNCDIF
                     VLPSYKEGFPVSVLEAKACGKVIVVSDCEGCVEAISNAYDGLWAKTKDSKDLIEKIQV
                     LLEDESLRINLGKNAAKDALQYDENVIAQRYLELYDRVIKNV"
     gene            1896. .2498
                     /gene="rfbP"
     CDS             1896. .2498
                     /gene="rfbP"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA62556.1"
                     /db_xref="GI:1050454"
                     /db_xref="SPTREMBL:P71124"
                     /translation="MYEKWIKRIFDFVLALFLLVLFSPLILITALLLKITQGSVIFTQ
                     NRPGLNEKIFKIYKFKTMSDERDEKGELLSDELRLKAFGKIVRSLSLDELLQLFNVLK
                     GDMSFVGPRPLLVEYLPLYNEEQKLRHKVRPGITGWAQVNGRNAISWQKKFELDVYYV
                     KNISFLLDLKIMFLTALKVLKRSGVSKEGHVTTEKFNGKN"
BASE COUNT      886 a    307 c    516 g    928 t
ORIGIN
```

```
  Query Match             62.1%;  Score 23.6;  DB 2;  Length 2637;
  Best Local Similarity   76.3%;  Pred. No. 3.6e+02;
  Matches   29;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38
          ||||||||  ||||  ||||||  ||||   |||||| |
Db   1384 AGGCTTTTCATCAATTTTTAAATTTTGCCAAAAAGCTT 1347
```

```
RESULT   7
AL139008
LOCUS       AL139008   159974 bp    DNA             PRI       24-JUL-2000
DEFINITION  Human DNA sequence from clone RP11-255A11 on chromosome 9. Contains
            the gene for a novel protein similar to suppressor of G2 allele of
            SKP1 homolog, the gene for a novel protein similar to KIAA1074, the
            gene for a novel protein similar to melanoma antigen, a putative
            novel gene, a pseudogene similar to Trypsin domain, a T cell
            receptor beta chain pseudogene, the gene for a novel protein
            similar to annexin A2 (ANXA2) (lipocortin II, calpactin I heavy
            chain, chromobindin 8, PAP-IV) and up to four putative genes for
            novel T cell receptor beta chain V region proteins. Contains ESTs,
```

STSs, GSSs and two CpG islands, complete sequence.
ACCESSION AL139008
VERSION AL139008.10 GI:8574139
KEYWORDS HTG; annexin; ANXA2; calpactin; chromobindin; CpG island; KIAA1074; lipocortin; melanoma; SKP1; T cell receptor; trypsin.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 159974)
AUTHORS Laird,G.
TITLE Direct Submission
JOURNAL Submitted (19-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Jun 20, 2000 this sequence version replaced gi:8247087.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr9
RP11-255A11 is from the library RPCI-11.1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/
VECTOR: pBACe3.6
This sequence is the entire insert of clone RP11-255A11 The true left end of clone RP11-384P7 is at 98021 in this sequence. The true right end of clone RP11-115O15 is at 30758 in this sequence.

FEATURES Location/Qualifiers

source	1. .159974
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="9"
	/clone="RP11-255A11"
	/clone_lib="RPCI-11.1"
misc_feature	824. .1219
	/note="match: GSS: Em:AQ060763"
repeat_region	839. .904
	/note="L2 repeat: matches 2677. .2744 of consensus"
misc_feature	843. .1196
	/note="match: GSS: Em:AQ131788"
repeat_region	1173. .1317
	/note="MIR repeat: matches 106. .262 of consensus"
misc_feature	1198. .1332
	/note="match: GSS: Em:AQ131788"
repeat_region	1425. .1487
	/note="MIR repeat: matches 68. .130 of consensus"
repeat_region	1665. .1796
	/note="MIR repeat: matches 90. .236 of consensus"
misc_feature	complement(2128. .2447)
	/note="match: GSS: Em:AQ564079"
misc_feature	2448. .2562
	/note="match: GSS: Em:AQ392027"
repeat_region	2479. .2757
	/note="AluSg repeat: matches 2. .293 of consensus"
misc_feature	2566. .2834
	/note="match: GSS: Em:AQ392027"
misc_feature	complement(2668. .3146)
	/note="match: GSS: Em:AQ473774"
misc_feature	complement(2729. .3157)
	/note="match: GSS: Em:AQ434620"
misc_feature	3176. .3363
	/note="match: GSS: Em:AQ476898"
misc_feature	3377. .3688
	/note="match: GSS: Em:AQ476898"
misc_feature	3536. .3681
	/note="match: GSS: Em:AQ372648"
misc_feature	3536. .3691
	/note="match: GSS: Em:AQ765454"
misc_feature	complement(3537. .3687)
	/note="match: GSS: Em:AZ023172"
misc_feature	complement(3541. .3688)
	/note="match: GSS: Em:AQ781128"
repeat_region	3617. .3682
	/note="33 copies 2 mer ta 90% conserved"
repeat_region	4079. .4376
	/note="AluSx repeat: matches 1. .298 of consensus"
repeat_region	4490. .4796
	/note="AluY repeat: matches 1. .307 of consensus"
repeat_region	5328. .5640
	/note="AluSq repeat: matches 1. .312 of consensus"
repeat_region	5985. .6294
	/note="AluSx repeat: matches 1. .311 of consensus"
repeat_region	6611. .6858
	/note="AluSx repeat: matches 40. .287 of consensus"
repeat_region	7139. .7433
	/note="AluSx repeat: matches 1. .292 of consensus"
repeat_region	7654. .7958
	/note="AluSx repeat: matches 1. .305 of consensus"
repeat_region	7999. .8155
	/note="FRAM repeat: matches 5. .161 of consensus"
misc_feature	complement(8224. .8804)
	/note="match: GSS: Em:AQ541215"
misc_feature	complement(8413. .8702)
	/note="match: GSS: Em:AQ093651"
mRNA	complement(join(8469. .8657,9352. .9405,13173. .13243, 13944. .14013,15434. .>15525))
	/gene="bA255A11.1"
	/note="match: cDNAs: Em:AF068289 Em:AF132856 match: ESTs: Em:AA416876 Em:AW376741 Em:AA305846 Em:AA312842 Em:AW505291 Em:AA470048 Em:AI479754 Em:AW363086 Em:AA935807"
	/evidence=not_experimental
	/product="bA255A11.1 (novel protein similar to suppressor of G2 allele of SKP1 homolog)"
polyA_site	complement(8469)
	/gene="bA255A11.1"
gene	complement(8469. .15525)
	/gene="bA255A11.1"
polyA_signal	complement(8485. .8492)
	/gene="bA255A11.1"
CDS	complement(join(8599. .8657,9352. .9405,13173. .13243, 13944. .14013,15434. .15494))
	/gene="bA255A11.1"
	/note="match: proteins: Tr:Q9Y2Z0"
	/codon_start=1
	/evidence=not_experimental
	/product="bA255A11.1 (novel protein similar to suppressor of G2 allele of SKP1 homolog)"
	/protein_id="CAB99337.1"
	/db_xref="GI:12314194"
	/translation="MMHIYYRQRAYCHILLGNYCVAVADAKKSREFNPNNSTAVLRKG ICEYHLKNYAAALETFIGGQKLVSADANFSDWIKRCQEAQNGSESEVVSPEFSFFMFL LT"
repeat_region	8833. .9124
	/note="AluSg repeat: matches 1. .291 of consensus"
repeat_region	9452. .9527
	/note="L2 repeat: matches 2665. .2740 of consensus"
misc_feature	10296. .10484
	/note="match: GSS: Em:B53940"

```
     repeat_region   10584. .10891
                     /note="AluSg repeat: matches 1. .307 of consensus"
     repeat_region   10969. .11286
                     /note="AluJo repeat: matches 1. .309 of consensus"
     repeat_region   11349. .11645
                     /note="AluJb repeat: matches 1. .295 of consensus"
     repeat_region   11964. .12183
                     /note="MER20 repeat: matches 1. .218 of consensus"
     repeat_region   12591. .12899
                     /note="AluSc repeat: matches 1. .309 of consensus"
     repeat_region   14038. .14348
                     /note="AluSg repeat: matches 1. .307 of consensus"
     repeat_region   14615. .14906
                     /note="AluSq repeat: matches 1. .293 of consensus"
     repeat_region   14978. .15101
                     /note="FLAM_C repeat: matches 6. .127 of consensus"
     misc_feature    complement(15363. .15659)
                     /note="match: GSS: Em:AQ154284"
     misc_feature    complement(15443. .15729)
                     /note="match: GSS: Em:B16076"
     repeat_region   16051. .16234
                     /note="L1M4 repeat: matches 4050. .4256 of consensus"
     repeat_region   16402. .16809
                     /note="LTR32 repeat: matches 14. .464 of consensus"
     repeat_region   16890. .16936
                     /note="L1ME repeat: matches 5369. .5416 of consensus"
     repeat_region   16937. .17233
                     /note="AluSx repeat: matches 1. .298 of consensus"
     repeat_region   17234. .17339
                     /note="L1ME repeat: matches 5416. .5517 of consensus"
     mRNA            complement(join(17773. .17906,19438. .>19587))
                     /gene="bA255A11.2"
                     /note="match: ESTs: Em:AA626249 Em:AI125587"
                     /evidence=not_experimental
                     /product="bA255A11.2 (putative novel protein)"
     gene            complement(join(17773. .17906,19438. .19587))
                     /gene="bA255A11.2"
     polyA_site      complement(17773)
                     /gene="bA255A11.2"
     polyA_site      complement(17777)
                     /gene="bA255A11.2"
     polyA_signal    complement(17785. .17792)
                     /gene="bA255A11.2"
     repeat_region   18875. .19162

  Query Match             62.1%;  Score 23.6;  DB 89;  Length 159974;
  Best Local Similarity   86.7%;  Pred. No. 1.6e+02;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       5 ttttgctcaaagtttaaactttgagcaaaa 34
           ||||||||| |||||  ||| |||||||||
Db   26194 TTTTGCTCAGAGTTTGTACTCTGAGCAAAA 26223


RESULT   8
AL139008/c
LOCUS       AL139008   159974 bp    DNA             PRI       24-JUL-2000
DEFINITION  Human DNA sequence from clone RP11-255A11 on chromosome 9. Contains
            the gene for a novel protein similar to suppressor of G2 allele of
            SKP1 homolog, the gene for a novel protein similar to KIAA1074, the
            gene for a novel protein similar to melanoma antigen, a putative
            novel gene, a pseudogene similar to Trypsin domain, a T cell
            receptor beta chain pseudogene, the gene for a novel protein
            similar to annexin A2 (ANXA2) (lipocortin II, calpactin I heavy
            chain, chromobindin 8, PAP-IV) and up to four putative genes for
            novel T cell receptor beta chain V region proteins. Contains ESTs,
            STSs, GSSs and two CpG islands, complete sequence.
ACCESSION   AL139008
VERSION     AL139008.10  GI:8574139
KEYWORDS    HTG; annexin; ANXA2; calpactin; chromobindin; CpG island; KIAA1074;
            lipocortin; melanoma; SKP1; T cell receptor; trypsin.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159974)
  AUTHORS   Laird,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 20, 2000 this sequence version replaced gi:8247087.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-255A11 is from the library RPCI-11.1 constructed at the
            Roswell Park Cancer Institute by the group of Pieter de Jong. For
            further details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-255A11 The true
            left end of clone RP11-384P7 is at 98021 in this sequence. The true
            right end of clone RP11-115O15 is at 30758 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .159974
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-255A11"
                     /clone_lib="RPCI-11.1"
     misc_feature    824. .1219
                     /note="match: GSS: Em:AQ060763"
     repeat_region   839. .904
                     /note="L2 repeat: matches 2677. .2744 of consensus"
     misc_feature    843. .1196
                     /note="match: GSS: Em:AQ131788"
     repeat_region   1173. .1317
                     /note="MIR repeat: matches 106. .262 of consensus"
     misc_feature    1198. .1332
                     /note="match: GSS: Em:AQ131788"
     repeat_region   1425. .1487
                     /note="MIR repeat: matches 68. .130 of consensus"
     repeat_region   1665. .1796
                     /note="MIR repeat: matches 90. .236 of consensus"
     misc_feature    complement(2128. .2447)
                     /note="match: GSS: Em:AQ564079"
     misc_feature    2448. .2562
                     /note="match: GSS: Em:AQ392027"
     repeat_region   2479. .2757
                     /note="AluSg repeat: matches 2. .293 of consensus"
     misc_feature    2566. .2834
                     /note="match: GSS: Em:AQ392027"
     misc_feature    complement(2668. .3146)
                     /note="match: GSS: Em:AQ473774"
     misc_feature    complement(2729. .3157)
                     /note="match: GSS: Em:AQ434620"
     misc_feature    3176. .3363
                     /note="match: GSS: Em:AQ476898"
     misc_feature    3377. .3688
```

```
                     /note="match: GSS: Em:AQ476898"
misc_feature         3536. .3681
                     /note="match: GSS: Em:AQ372648"
misc_feature         3536. .3691
                     /note="match: GSS: Em:AQ765454"
misc_feature         complement(3537. .3687)
                     /note="match: GSS: Em:AZ023172"
misc_feature         complement(3541. .3688)
                     /note="match: GSS: Em:AQ781128"
repeat_region        3617. .3682
                     /note="33 copies 2 mer ta 90% conserved"
repeat_region        4079. .4376
                     /note="AluSx repeat: matches 1. .298 of consensus"
repeat_region        4490. .4796
                     /note="AluY repeat: matches 1. .307 of consensus"
repeat_region        5328. .5640
                     /note="AluSq repeat: matches 1. .312 of consensus"
repeat_region        5985. .6294
                     /note="AluSx repeat: matches 1. .311 of consensus"
repeat_region        6611. .6858
                     /note="AluSx repeat: matches 40. .287 of consensus"
repeat_region        7139. .7433
                     /note="AluSx repeat: matches 1. .292 of consensus"
repeat_region        7654. .7958
                     /note="AluSx repeat: matches 1. .305 of consensus"
repeat_region        7999. .8155
                     /note="FRAM repeat: matches 5. .161 of consensus"
misc_feature         complement(8224. .8804)
                     /note="match: GSS: Em:AQ541215"
misc_feature         complement(8413. .8702)
                     /note="match: GSS: Em:AQ093651"
mRNA                 complement(join(8469. .8657,9352. .9405,13173. .13243,
                     13944. .14013,15434. .>15525))
                     /gene="bA255A11.1"
                     /note="match: cDNAs: Em:AF068289 Em:AF132856
                     match: ESTs: Em:AA416876 Em:AW376741 Em:AA305846
                     Em:AA312842 Em:AW505291 Em:AA470048 Em:AI479754
                     Em:AW363086 Em:AA935807"
                     /evidence=not_experimental
                     /product="bA255A11.1 (novel protein similar to suppressor
                     of G2 allele of SKP1 homolog)"
polyA_site           complement(8469)
                     /gene="bA255A11.1"
gene                 complement(8469. .15525)
                     /gene="bA255A11.1"
polyA_signal         complement(8485. .8492)
                     /gene="bA255A11.1"
CDS                  complement(join(8599. .8657,9352. .9405,13173. .13243,
                     13944. .14013,15434. .15494))
                     /gene="bA255A11.1"
                     /note="match: proteins: Tr:Q9Y2Z0"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="bA255A11.1 (novel protein similar to suppressor
                     of G2 allele of SKP1 homolog)"
                     /protein_id="CAB99337.1"
                     /db_xref="GI:12314194"
                     /translation="MMHIYYRQRAYCHILLGNYCVAVADAKKSREFNPNNSTAVLRKG
                     ICEYHLKNYAAALETFIGGQKLVSADANFSDWIKRCQEAQNGSESEVVSPEFSFFMFL
                     LT"
repeat_region        8833. .9124
                     /note="AluSg repeat: matches 1. .291 of consensus"
repeat_region        9452. .9527
                     /note="L2 repeat: matches 2665. .2740 of consensus"
misc_feature         10296. .10484
                     /note="match: GSS: Em:B53940"
repeat_region        10584. .10891
                     /note="AluSg repeat: matches 1. .307 of consensus"
repeat_region        10969. .11286
                     /note="AluJo repeat: matches 1. .309 of consensus"
repeat_region        11349. .11645
                     /note="AluJb repeat: matches 1. .295 of consensus"
repeat_region        11964. .12183
                     /note="MER20 repeat: matches 1. .218 of consensus"
repeat_region        12591. .12899
                     /note="AluSc repeat: matches 1. .309 of consensus"
repeat_region        14038. .14348
                     /note="AluSg repeat: matches 1. .307 of consensus"
repeat_region        14615. .14906
                     /note="AluSq repeat: matches 1. .293 of consensus"
repeat_region        14978. .15101
                     /note="FLAM_C repeat: matches 6. .127 of consensus"
misc_feature         complement(15363. .15659)
                     /note="match: GSS: Em:AQ154284"
misc_feature         complement(15443. .15729)
                     /note="match: GSS: Em:B16076"
repeat_region        16051. .16234
                     /note="L1M4 repeat: matches 4050. .4256 of consensus"
repeat_region        16402. .16809
                     /note="LTR32 repeat: matches 14. .464 of consensus"
repeat_region        16890. .16936
                     /note="L1ME repeat: matches 5369. .5416 of consensus"
repeat_region        16937. .17233
                     /note="AluSx repeat: matches 1. .298 of consensus"
repeat_region        17234. .17339
                     /note="L1ME repeat: matches 5416. .5517 of consensus"
mRNA                 complement(join(17773. .17906,19438. .>19587))
                     /gene="bA255A11.2"
                     /note="match: ESTs: Em:AA626249 Em:AI125587"
                     /evidence=not_experimental
                     /product="bA255A11.2 (putative novel protein)"
gene                 complement(join(17773. .17906,19438. .19587))
                     /gene="bA255A11.2"
polyA_site           complement(17773)
                     /gene="bA255A11.2"
polyA_site           complement(17777)
                     /gene="bA255A11.2"
polyA_signal         complement(17785. .17792)
                     /gene="bA255A11.2"
repeat_region        18875. .19162

  Query Match             62.1%;  Score 23.6;  DB 89;  Length 159974;
  Best Local Similarity   86.7%;  Pred. No. 1.6e+02;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       5 ttttgctcaaagtttaaactttgagcaaaa 34
            ||||||||| |||  ||||| |||||||||
Db   26223 TTTTGCTCAGAGTACAAACTCTGAGCAAAA 26194


RESULT   9
AL353675
LOCUS       AL353675   169223 bp    DNA             HTG       05-SEP-2000
DEFINITION  Homo sapiens chromosome 9 clone RP11-115O15, *** SEQUENCING IN
            PROGRESS ***, 9 unordered pieces.
ACCESSION   AL353675
VERSION     AL353675.12  GI:9988294
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169223)
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 6, 2000 this sequence version replaced gi:9801707.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA115O15
```

```
            ---------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 166896 bases at least Q40
            Consensus quality: 167535 bases at least Q30
            Consensus quality: 167945 bases at least Q20
            Insert size: 168423; sum-of-contigs
            Insert size: 168974; 4.4% error; agarose-fp
            Quality coverage: 6.75x in Q20 bases; sum-of-contigs Quality
            coverage: 6.73x in Q20 bases; agarose-fp
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    39170: contig of 39170 bp in length
            *    39171 39270: gap of      100 bp
            *    39271    43291: contig of 4021 bp in length
            *    43292 43391: gap of      100 bp
            *    43392    75740: contig of 32349 bp in length
            *    75741 75840: gap of      100 bp
            *    75841   104324: contig of 28484 bp in length
            *   104325 104424: gap of      100 bp
            *   104425   112491: contig of 8067 bp in length
            *   112492 112591: gap of      100 bp
            *   112592   116081: contig of 3490 bp in length
            *   116082 116181: gap of      100 bp
            *   116182   119277: contig of 3096 bp in length
            *   119278 119377: gap of      100 bp
            *   119378   127902: contig of 8525 bp in length
            *   127903 128002: gap of      100 bp
            *   128003   169223: contig of 41221 bp in length.
FEATURES             Location/Qualifiers
     source          1. .169223
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-115O15"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .39170
                     /note="assembly_fragment:02240
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    39271. .43291
                     /note="assembly_fragment:01832
                     fragment_chain:1"
     misc_feature    43392. .75740
                     /note="assembly_fragment:01323
                     fragment_chain:1"
     misc_feature    75841. .104324
                     /note="assembly_fragment:02411
                     fragment_chain:1"
     misc_feature    104425. .112491
                     /note="assembly_fragment:03019
                     fragment_chain:1"
     misc_feature    112592. .116081
                     /note="assembly_fragment:00606
                     fragment_chain:1"
     misc_feature    116182. .119277
                     /note="assembly_fragment:00279"
     misc_feature    119378. .127902
                     /note="assembly_fragment:01929
                     fragment_chain:2"
     misc_feature    128003. .169223
                     /note="assembly_fragment:01181
                     fragment_chain:2
                     clone_end:T7
                     vector_side:right"
BASE COUNT    41936 a  40234 c  41157 g  45095 t    801 others
ORIGIN


  Query Match             62.1%;  Score 23.6;  DB 79;  Length 169223;
  Best Local Similarity   86.7%;  Pred. No. 1.5e+02;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       5 ttttgctcaaagtttaaactttgagcaaaa 34
           ||||||||| |||  ||||| |||||||||
Db    4530 TTTTGCTCAGAGTACAAACTCTGAGCAAAA 4559


RESULT  10
AL353675/c
LOCUS       AL353675   169223 bp    DNA             HTG       05-SEP-2000
DEFINITION  Homo sapiens chromosome 9 clone RP11-115O15, *** SEQUENCING IN
            PROGRESS ***, 9 unordered pieces.
ACCESSION   AL353675
VERSION     AL353675.12  GI:9988294
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169223)
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 6, 2000 this sequence version replaced gi:9801707.
            ---------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ---------- Project Information
            Center project name: bA115O15
            ---------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 166896 bases at least Q40
            Consensus quality: 167535 bases at least Q30
            Consensus quality: 167945 bases at least Q20
            Insert size: 168423; sum-of-contigs
            Insert size: 168974; 4.4% error; agarose-fp
            Quality coverage: 6.75x in Q20 bases; sum-of-contigs Quality
            coverage: 6.73x in Q20 bases; agarose-fp
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    39170: contig of 39170 bp in length
            *    39171 39270: gap of      100 bp
            *    39271    43291: contig of 4021 bp in length
            *    43292 43391: gap of      100 bp
            *    43392    75740: contig of 32349 bp in length
            *    75741 75840: gap of      100 bp
            *    75841   104324: contig of 28484 bp in length
            *   104325 104424: gap of      100 bp
            *   104425   112491: contig of 8067 bp in length
            *   112492 112591: gap of      100 bp
            *   112592   116081: contig of 3490 bp in length
            *   116082 116181: gap of      100 bp
            *   116182   119277: contig of 3096 bp in length
```

```
          *   119278 119377: gap of      100 bp
          *   119378   127902: contig of 8525 bp in length
          *   127903 128002: gap of      100 bp
          *   128003   169223: contig of 41221 bp in length.
FEATURES             Location/Qualifiers
     source          1. .169223
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-115O15"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .39170
                     /note="assembly_fragment:02240
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    39271. .43291
                     /note="assembly_fragment:01832
                     fragment_chain:1"
     misc_feature    43392. .75740
                     /note="assembly_fragment:01323
                     fragment_chain:1"
     misc_feature    75841. .104324
                     /note="assembly_fragment:02411
                     fragment_chain:1"
     misc_feature    104425. .112491
                     /note="assembly_fragment:03019
                     fragment_chain:1"
     misc_feature    112592. .116081
                     /note="assembly_fragment:00606
                     fragment_chain:1"
     misc_feature    116182. .119277
                     /note="assembly_fragment:00279"
     misc_feature    119378. .127902
                     /note="assembly_fragment:01929
                     fragment_chain:2"
     misc_feature    128003. .169223
                     /note="assembly_fragment:01181
                     fragment_chain:2
                     clone_end:T7
                     vector_side:right"
BASE COUNT    41936 a  40234 c  41157 g  45095 t    801 others
ORIGIN


  Query Match             62.1%;  Score 23.6;  DB 79;  Length 169223;
  Best Local Similarity   86.7%;  Pred. No. 1.5e+02;
  Matches   26;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       5 ttttgctcaaagtttaaactttgagcaaaa 34
           ||||||||| |||||  ||| |||||||||
Db    4559 TTTTGCTCAGAGTTTGTACTCTGAGCAAAA 4530


RESULT  11
SYNFLUAPB
LOCUS       SYNFLUAPB     164 bp    DNA             SYN       27-APR-1993
DEFINITION  Plasmid vector DNA and Influenza A virus promoter.
ACCESSION   M25075
VERSION     M25075.1  GI:208383
KEYWORDS    .
SOURCE      Plasmid SV2cat DNA and Influenza virus cDNA to RNA.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 164)
  AUTHORS   Yamanaka,K., Ishihama,A. and Nagata,K.
  TITLE     Translational regulation of Influenza virus mRNAs
  JOURNAL   Virus Genes 2, 19-30 (1988)
  MEDLINE   89147186
FEATURES             Location/Qualifiers
     source          1. .164
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT       51 a     28 c     47 g     38 t
ORIGIN


  Query Match             61.1%;  Score 23.2;  DB 56;  Length 164;
  Best Local Similarity   77.8%;  Pred. No. 8.5e+02;
  Matches   28;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 aggcttttgctcaaagtttaaactttgagcaaaagc 36
           ||||||||||  |||| ||  | ||  |||||||||
Db      49 AGGCTTTTGCAAAAAGCTTGGAATTCCAGCAAAAGC 84


RESULT  12
SYNFLUAPB/c
LOCUS       SYNFLUAPB     164 bp    DNA             SYN       27-APR-1993
DEFINITION  Plasmid vector DNA and Influenza A virus promoter.
ACCESSION   M25075
VERSION     M25075.1  GI:208383
KEYWORDS    .
SOURCE      Plasmid SV2cat DNA and Influenza virus cDNA to RNA.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 164)
  AUTHORS   Yamanaka,K., Ishihama,A. and Nagata,K.
  TITLE     Translational regulation of Influenza virus mRNAs
  JOURNAL   Virus Genes 2, 19-30 (1988)
  MEDLINE   89147186
FEATURES             Location/Qualifiers
     source          1. .164
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT       51 a     28 c     47 g     38 t
ORIGIN


  Query Match             61.1%;  Score 23.2;  DB 56;  Length 164;
  Best Local Similarity   77.8%;  Pred. No. 8.5e+02;
  Matches   28;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       3 gcttttgctcaaagtttaaactttgagcaaaagcct 38
           |||||||||  || |  || ||||  ||||||||||
Db      84 GCTTTTGCTGGAATTCCAAGCTTTTTGCAAAAGCCT 49


RESULT  13
AC021854
LOCUS       AC021854   144402 bp    DNA             HTG       07-JUL-2000
DEFINITION  Homo sapiens chromosome 5 clone RP11-505H15, WORKING DRAFT
            SEQUENCE, 21 unordered pieces.
ACCESSION   AC021854
VERSION     AC021854.3  GI:7658486
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 144402)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 144402)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Apr 28, 2000 this sequence version replaced gi:7024086.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
```

```
            ---------- Project Information ----------
            Center project name: H_NH0505H15
            ---------- Summary Statistics  ----------
            Sequencing vector: M13; 72%
            Sequencing vector: plasmid; 28%
            Chemistry: Dye-primer ET; 72% of reads
            Chemistry: Dye-terminator Big Dye; 28% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 131476 bases at least Q40
            Consensus quality: 134731 bases at least Q30
            Consensus quality: 137492 bases at least Q20
            Insert size: 156000; agarose-fp
            Insert size: 142402; sum-of-contigs
            Quality coverage: 3.42 in Q20 bases; agarose-fp
            Quality coverage: 3.82 in Q20 bases; sum-of-contigs
            ----------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1445: contig of 1445 bp in length
            *     1446     1545: gap of unknown length
            *     1546     3851: contig of 2306 bp in length
            *     3852     3951: gap of unknown length
            *     3952     6428: contig of 2477 bp in length
            *     6429     6528: gap of unknown length
            *     6529     7846: contig of 1318 bp in length
            *     7847     7946: gap of unknown length
            *     7947    10364: contig of 2418 bp in length
            *    10365    10464: gap of unknown length
            *    10465    13198: contig of 2734 bp in length
            *    13199    13298: gap of unknown length
            *    13299    15320: contig of 2022 bp in length
            *    15321    15420: gap of unknown length
            *    15421    18740: contig of 3320 bp in length
            *    18741    18840: gap of unknown length
            *    18841    22611: contig of 3771 bp in length
            *    22612    22711: gap of unknown length
            *    22712    26922: contig of 4211 bp in length
            *    26923    27022: gap of unknown length
            *    27023    32530: contig of 5508 bp in length
            *    32531    32630: gap of unknown length
            *    32631    39868: contig of 7238 bp in length
            *    39869    39968: gap of unknown length
            *    39969    47269: contig of 7301 bp in length
            *    47270    47369: gap of unknown length
            *    47370    55269: contig of 7900 bp in length
            *    55270    55369: gap of unknown length
            *    55370    62905: contig of 7536 bp in length
            *    62906    63005: gap of unknown length
            *    63006    70907: contig of 7902 bp in length
            *    70908    71007: gap of unknown length
            *    71008    81750: contig of 10743 bp in length
            *    81751    81850: gap of unknown length
            *    81851    96456: contig of 14606 bp in length
            *    96457    96556: gap of unknown length
            *    96557   112579: contig of 16023 bp in length
            *   112580   112679: gap of unknown length
            *   112680   129486: contig of 16807 bp in length
            *   129487   129586: gap of unknown length
            *   129587   144402: contig of 14816 bp in length.
FEATURES             Location/Qualifiers
     source          1. .144402
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-505H15"
     misc_feature    1. .1445
                     /note="assembly_name:Contig10"
     misc_feature    1546. .3851
                     /note="assembly_name:Contig11"
     misc_feature    3952. .6428
                     /note="assembly_name:Contig12"
     misc_feature    6529. .7846
                     /note="assembly_name:Contig13
                     clone_end:T7
                     vector_side:left"
     misc_feature    7947. .10364
                     /note="assembly_name:Contig14"
     misc_feature    10465. .13198
                     /note="assembly_name:Contig15"
     misc_feature    13299. .15320
                     /note="assembly_name:Contig16"
     misc_feature    15421. .18740
                     /note="assembly_name:Contig17"
     misc_feature    18841. .22611
                     /note="assembly_name:Contig18"
     misc_feature    22712. .26922
                     /note="assembly_name:Contig19"
     misc_feature    27023. .32530
                     /note="assembly_name:Contig20"
     misc_feature    32631. .39868
                     /note="assembly_name:Contig21"
     misc_feature    39969. .47269
                     /note="assembly_name:Contig22"
     misc_feature    47370. .55269
                     /note="assembly_name:Contig23
                     clone_end:SP6
                     vector_side:left"
     misc_feature    55370. .62905
                     /note="assembly_name:Contig24"
     misc_feature    63006. .70907
                     /note="assembly_name:Contig25"
     misc_feature    71008. .81750
                     /note="assembly_name:Contig26"
     misc_feature    81851. .96456
                     /note="assembly_name:Contig27"
     misc_feature    96557. .112579
                     /note="assembly_name:Contig28"
     misc_feature    112680. .129486
                     /note="assembly_name:Contig29"
     misc_feature    129587. .144402
                     /note="assembly_name:Contig30"
BASE COUNT    40809 a  30400 c  30521 g  40644 t   2028 others
ORIGIN


  Query Match             60.0%;  Score 22.8;  DB 67;  Length 144402;
  Best Local Similarity   79.4%;  Pred. No. 2.9e+02;
  Matches   27;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 cttttgctcaaagtttaaactttgagcaaaagcc 37
           |||  ||||| |||||||| ||| || ||||| |
Db   78263 CTTAGGCTCACAGTTTAAAATTTAAGAAAAAGAC 78296


RESULT  14
AC021854/c
LOCUS       AC021854   144402 bp    DNA             HTG       07-JUL-2000
DEFINITION  Homo sapiens chromosome 5 clone RP11-505H15, WORKING DRAFT
            SEQUENCE, 21 unordered pieces.
ACCESSION   AC021854
VERSION     AC021854.3  GI:7658486
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 144402)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 144402)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Apr 28, 2000 this sequence version replaced gi:7024086.

            ---------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0505H15
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 72%
            Sequencing vector: plasmid; 28%
            Chemistry: Dye-primer ET; 72% of reads
            Chemistry: Dye-terminator Big Dye; 28% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 131476 bases at least Q40
            Consensus quality: 134731 bases at least Q30
            Consensus quality: 137492 bases at least Q20
            Insert size: 156000; agarose-fp
            Insert size: 142402; sum-of-contigs
            Quality coverage: 3.42 in Q20 bases; agarose-fp
            Quality coverage: 3.82 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1445: contig of 1445 bp in length
            *     1446     1545: gap of unknown length
            *     1546     3851: contig of 2306 bp in length
            *     3852     3951: gap of unknown length
            *     3952     6428: contig of 2477 bp in length
            *     6429     6528: gap of unknown length
            *     6529     7846: contig of 1318 bp in length
            *     7847     7946: gap of unknown length
            *     7947    10364: contig of 2418 bp in length
            *    10365    10464: gap of unknown length
            *    10465    13198: contig of 2734 bp in length
            *    13199    13298: gap of unknown length
            *    13299    15320: contig of 2022 bp in length
            *    15321    15420: gap of unknown length
            *    15421    18740: contig of 3320 bp in length
            *    18741    18840: gap of unknown length
            *    18841    22611: contig of 3771 bp in length
            *    22612    22711: gap of unknown length
            *    22712    26922: contig of 4211 bp in length
            *    26923    27022: gap of unknown length
            *    27023    32530: contig of 5508 bp in length
            *    32531    32630: gap of unknown length
            *    32631    39868: contig of 7238 bp in length
            *    39869    39968: gap of unknown length
            *    39969    47269: contig of 7301 bp in length
            *    47270    47369: gap of unknown length
            *    47370    55269: contig of 7900 bp in length
            *    55270    55369: gap of unknown length
            *    55370    62905: contig of 7536 bp in length
            *    62906    63005: gap of unknown length
            *    63006    70907: contig of 7902 bp in length
            *    70908    71007: gap of unknown length
            *    71008    81750: contig of 10743 bp in length
            *    81751    81850: gap of unknown length
            *    81851    96456: contig of 14606 bp in length
            *    96457    96556: gap of unknown length
            *    96557   112579: contig of 16023 bp in length
            *   112580   112679: gap of unknown length
            *   112680   129486: contig of 16807 bp in length
            *   129487   129586: gap of unknown length
            *   129587   144402: contig of 14816 bp in length.
FEATURES             Location/Qualifiers
     source          1. .144402
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-505H15"
     misc_feature    1. .1445
                     /note="assembly_name:Contig10"
     misc_feature    1546. .3851
                     /note="assembly_name:Contig11"
     misc_feature    3952. .6428
                     /note="assembly_name:Contig12"
     misc_feature    6529. .7846
                     /note="assembly_name:Contig13
                     clone_end:T7
                     vector_side:left"
     misc_feature    7947. .10364
                     /note="assembly_name:Contig14"
     misc_feature    10465. .13198
                     /note="assembly_name:Contig15"
     misc_feature    13299. .15320
                     /note="assembly_name:Contig16"
     misc_feature    15421. .18740
                     /note="assembly_name:Contig17"
     misc_feature    18841. .22611
                     /note="assembly_name:Contig18"
     misc_feature    22712. .26922
                     /note="assembly_name:Contig19"
     misc_feature    27023. .32530
                     /note="assembly_name:Contig20"
     misc_feature    32631. .39868
                     /note="assembly_name:Contig21"
     misc_feature    39969. .47269
                     /note="assembly_name:Contig22"
     misc_feature    47370. .55269
                     /note="assembly_name:Contig23
                     clone_end:SP6
                     vector_side:left"
     misc_feature    55370. .62905
                     /note="assembly_name:Contig24"
     misc_feature    63006. .70907
                     /note="assembly_name:Contig25"
     misc_feature    71008. .81750
                     /note="assembly_name:Contig26"
     misc_feature    81851. .96456
                     /note="assembly_name:Contig27"
     misc_feature    96557. .112579
                     /note="assembly_name:Contig28"
     misc_feature    112680. .129486
                     /note="assembly_name:Contig29"
     misc_feature    129587. .144402
                     /note="assembly_name:Contig30"
BASE COUNT    40809 a  30400 c  30521 g  40644 t   2028 others
ORIGIN


  Query Match             60.0%;  Score 22.8;  DB 67;  Length 144402;
  Best Local Similarity   79.4%;  Pred. No. 2.9e+02;
  Matches   27;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      2 ggcttttgctcaaagtttaaactttgagcaaaag 35
          | ||||| || ||| |||||||| ||||| |||
Db  78296 GTCTTTTTCTTAAATTTTAAACTGTGAGCCTAAG 78263


RESULT  15
AC024662
LOCUS       AC024662   171549 bp    DNA             HTG       12-MAY-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-355H11 map 4, WORKING DRAFT
```

```
            SEQUENCE, 19 unordered pieces.
ACCESSION   AC024662
VERSION     AC024662.3  GI:7770486
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171549)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 4, clone RP11-355H11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171549)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 12, 2000 this sequence version replaced gi:7523776.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L7315
                Center clone name: 355_H_11
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 160958 bases at least Q40
                Consensus quality: 165963 bases at least Q30
                Consensus quality: 168250 bases at least Q20
                Insert size: 188000; agarose-fp
                Insert size: 169749; sum-of-contigs
                Quality coverage: 4.1 in Q20 bases; agarose-fp
                Quality coverage: 4.5 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1045: contig of 1045 bp in length
            *     1046 1145: gap of      100 bp
            *     1146     1461: contig of 316 bp in length
            *     1462 1561: gap of      100 bp
            *     1562     3812: contig of 2251 bp in length
            *     3813 3912: gap of      100 bp
            *     3913     6193: contig of 2281 bp in length
            *     6194 6293: gap of      100 bp
            *     6294     9170: contig of 2877 bp in length
            *     9171 9270: gap of      100 bp
            *     9271    11732: contig of 2462 bp in length
            *    11733 11832: gap of      100 bp
            *    11833    14652: contig of 2820 bp in length
            *    14653 14752: gap of      100 bp
            *    14753    19897: contig of 5145 bp in length
            *    19898 19997: gap of      100 bp
            *    19998    25161: contig of 5164 bp in length
            *    25162 25261: gap of      100 bp
            *    25262    30394: contig of 5133 bp in length
            *    30395 30494: gap of      100 bp
            *    30495    39640: contig of 9146 bp in length
            *    39641 39740: gap of      100 bp
            *    39741    49008: contig of 9268 bp in length
            *    49009 49108: gap of      100 bp
            *    49109    59039: contig of 9931 bp in length
            *    59040 59139: gap of      100 bp
            *    59140    70556: contig of 11417 bp in length
            *    70557 70656: gap of      100 bp
            *    70657    83397: contig of 12741 bp in length
            *    83398 83497: gap of      100 bp
            *    83498    99129: contig of 15632 bp in length
            *    99130 99229: gap of      100 bp
            *    99230   117742: contig of 18513 bp in length
            *   117743 117842: gap of      100 bp
            *   117843   139073: contig of 21231 bp in length
            *   139074 139173: gap of      100 bp
            *   139174   171549: contig of 32376 bp in length.
FEATURES             Location/Qualifiers
     source          1. .171549
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-355H11"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1045
                     /note="assembly_fragment"
     misc_feature    1146. .1461
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    1562. .3812
                     /note="assembly_fragment"
     misc_feature    3913. .6193
                     /note="assembly_fragment"
     misc_feature    6294. .9170
                     /note="assembly_fragment"
     misc_feature    9271. .11732
                     /note="assembly_fragment"
     misc_feature    11833. .14652
                     /note="assembly_fragment"
     misc_feature    14753. .19897
                     /note="assembly_fragment"
     misc_feature    19998. .25161
                     /note="assembly_fragment"
     misc_feature    25262. .30394
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    30495. .39640
                     /note="assembly_fragment"
     misc_feature    39741. .49008
                     /note="assembly_fragment"
     misc_feature    49109. .59039
                     /note="assembly_fragment"
     misc_feature    59140. .70556
                     /note="assembly_fragment"
```

```
     misc_feature    70657. .83397
                     /note="assembly_fragment"
     misc_feature    83498. .99129
                     /note="assembly_fragment"
     misc_feature    99230. .117742
                     /note="assembly_fragment"
     misc_feature    117843. .139073
                     /note="assembly_fragment"
     misc_feature    139174. .171549
                     /note="assembly_fragment"
BASE COUNT    54897 a  29922 c  29937 g  54993 t   1800 others
ORIGIN


  Query Match             60.0%;  Score 22.8;  DB 69;  Length 171549;
  Best Local Similarity   79.4%;  Pred. No. 2.8e+02;
  Matches   27;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaa 34
          |||||| || |||||  |  ||||||||||| |
Db 166077 AGGCTTCTGATCAAACATACAACTTTGAGCAACA 166110


RESULT  16
AC024662/c
LOCUS       AC024662   171549 bp    DNA             HTG       12-MAY-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-355H11 map 4, WORKING DRAFT
            SEQUENCE, 19 unordered pieces.
ACCESSION   AC024662
VERSION     AC024662.3  GI:7770486
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171549)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 4, clone RP11-355H11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171549)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 12, 2000 this sequence version replaced gi:7523776.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L7315
                Center clone name: 355_H_11
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 160958 bases at least Q40
                Consensus quality: 165963 bases at least Q30
                Consensus quality: 168250 bases at least Q20
                Insert size: 188000; agarose-fp
                Insert size: 169749; sum-of-contigs
                Quality coverage: 4.1 in Q20 bases; agarose-fp
                Quality coverage: 4.5 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1045: contig of 1045 bp in length
            *    1046  1145: gap of      100 bp
            *    1146     1461: contig of 316 bp in length
            *    1462  1561: gap of      100 bp
            *    1562     3812: contig of 2251 bp in length
            *    3813  3912: gap of      100 bp
            *    3913     6193: contig of 2281 bp in length
            *    6194  6293: gap of      100 bp
            *    6294     9170: contig of 2877 bp in length
            *    9171  9270: gap of      100 bp
            *    9271    11732: contig of 2462 bp in length
            *   11733 11832: gap of      100 bp
            *   11833    14652: contig of 2820 bp in length
            *   14653 14752: gap of      100 bp
            *   14753    19897: contig of 5145 bp in length
            *   19898 19997: gap of      100 bp
            *   19998    25161: contig of 5164 bp in length
            *   25162 25261: gap of      100 bp
            *   25262    30394: contig of 5133 bp in length
            *   30395 30494: gap of      100 bp
            *   30495    39640: contig of 9146 bp in length
            *   39641 39740: gap of      100 bp
            *   39741    49008: contig of 9268 bp in length
            *   49009 49108: gap of      100 bp
            *   49109    59039: contig of 9931 bp in length
            *   59040 59139: gap of      100 bp
            *   59140    70556: contig of 11417 bp in length
            *   70557 70656: gap of      100 bp
            *   70657    83397: contig of 12741 bp in length
            *   83398 83497: gap of      100 bp
            *   83498    99129: contig of 15632 bp in length
            *   99130 99229: gap of      100 bp
            *   99230   117742: contig of 18513 bp in length
            *  117743 117842: gap of      100 bp
            *  117843   139073: contig of 21231 bp in length
            *  139074 139173: gap of      100 bp
            *  139174   171549: contig of 32376 bp in length.
FEATURES             Location/Qualifiers
     source          1. .171549
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-355H11"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1045
                     /note="assembly_fragment"
     misc_feature    1146. .1461
                     /note="assembly_fragment
                     clone_end:T7
```

```
                     vector_side:right"
     misc_feature    1562. .3812
                     /note="assembly_fragment"
     misc_feature    3913. .6193
                     /note="assembly_fragment"
     misc_feature    6294. .9170
                     /note="assembly_fragment"
     misc_feature    9271. .11732
                     /note="assembly_fragment"
     misc_feature    11833. .14652
                     /note="assembly_fragment"
     misc_feature    14753. .19897
                     /note="assembly_fragment"
     misc_feature    19998. .25161
                     /note="assembly_fragment"
     misc_feature    25262. .30394
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    30495. .39640
                     /note="assembly_fragment"
     misc_feature    39741. .49008
                     /note="assembly_fragment"
     misc_feature    49109. .59039
                     /note="assembly_fragment"
     misc_feature    59140. .70556
                     /note="assembly_fragment"
     misc_feature    70657. .83397
                     /note="assembly_fragment"
     misc_feature    83498. .99129
                     /note="assembly_fragment"
     misc_feature    99230. .117742
                     /note="assembly_fragment"
     misc_feature    117843. .139073
                     /note="assembly_fragment"
     misc_feature    139174. .171549
                     /note="assembly_fragment"
BASE COUNT    54897 a  29922 c  29937 g  54993 t   1800 others
ORIGIN


  Query Match             60.0%;  Score 22.8;  DB 69;  Length 171549;
  Best Local Similarity   79.4%;  Pred. No. 2.8e+02;
  Matches   27;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       5 ttttgctcaaagtttaaactttgagcaaaagcct 38
           | ||||||||||||  |  ||||| || ||||||
Db 166110 TGTTGCTCAAAGTTGTATGTTTGATCAGAAGCCT 166077


RESULT  17
AC019179
LOCUS       AC019179   196721 bp    DNA             PRI       08-NOV-2000
DEFINITION  Homo sapiens BAC clone RP11-240A16 from 4, complete sequence.
ACCESSION   AC019179
VERSION     AC019179.4  GI:11120947
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 196721)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 196721)
  AUTHORS   Harkins,R., Maupin,R., Gregory,S., Coblitz,B. and Fleming,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-240A16
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 196721)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 196721)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Nov 8, 2000 this sequence version replaced gi:7630907.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0240A16
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            Actual start of this clone is at base position 1 of RP11-240A16;
            actual end is at base position 196721 of RP11-240A16.

            The sequence H_NH0240A16 from base positon 157677 to 158503
            contains a tandem repeat.  The assembly is consistant with digest
            informationbut the sequence fidelity cannot be guaranteed.
FEATURES                Location/Qualifiers
     source             1. .196721
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="4"
                        /map="4"
                        /clone="RP11-240A16"
                        /clone_lib="RPCI-11"
     repeat_region      3448. .3840
                        /rpt_family="ERV"
     repeat_region      4103. .4596
                        /rpt_family="ERVK"
     repeat_region      5164. .5231
                        /rpt_family="MIR"
     repeat_region      5616. .5734
                        /rpt_family="MIR"
     repeat_region      5861. .5911
                        /rpt_family="Mariner"
     repeat_region      6619. .6749
```

```
                /rpt_family="MaLR"
repeat_region   6727. .6776
                /rpt_family="MaLR"
repeat_region   8458. .8684
                /rpt_family="Alu"
repeat_region   8685. .8843
                /rpt_family="Alu"
repeat_region   10345. .10912
                /rpt_family="MaLR"
repeat_region   10952. .11592
                /rpt_family="L2"
repeat_region   14824. .15153
                /rpt_family="L1"
repeat_region   16195. .16619
                /rpt_family="ERVL"
repeat_region   16649. .16822
                /rpt_family="MaLR"
repeat_region   17787. .17850
                /rpt_family="Alu"
repeat_region   18581. .18773
                /rpt_family="MER1_type"
repeat_region   19637. .19697
                /rpt_family="L2"
repeat_region   19798. .19971
                /rpt_family="Alu"
repeat_region   19991. .20128
                /rpt_family="Alu"
repeat_region   22389. .22580
                /rpt_family="MIR"
repeat_region   23974. .24287
                /rpt_family="Alu"
repeat_region   25165. .25399
                /rpt_family="MIR"
repeat_region   25434. .25719
                /rpt_family="Alu"
repeat_region   26014. .26157
                /rpt_family="L2"
repeat_region   28448. .28553
                /rpt_family="MIR"
repeat_region   29093. .29243
                /rpt_family="Alu"
repeat_region   29715. .30240
                /rpt_family="MaLR"
repeat_region   30717. .30823
                /rpt_family="L1"
repeat_region   30859. .30929
                /rpt_family="L1"
repeat_region   30935. .31516
                /rpt_family="ERV1"
repeat_region   31502. .31764
                /rpt_family="ERV1"
repeat_region   31871. .31981
                /rpt_family="L1"
repeat_region   31982. .32494
                /rpt_family="ERVK"
repeat_region   32495. .32523
                /rpt_family="L1"
repeat_region   32579. .32646
                /rpt_family="MaLR"
repeat_region   32775. .32922
                /rpt_family="ERVL"
repeat_region   33108. .33153
                /rpt_family="ERVL"
repeat_region   33187. .33486
                /rpt_family="Alu"
repeat_region   33588. .33696
                /rpt_family="MaLR"
repeat_region   34154. .34461
                /rpt_family="L2"
repeat_region   34462. .35265
                /rpt_family="ERV1"
repeat_region   35266. .35432
                /rpt_family="L2"
repeat_region   35517. .35685
                /rpt_family="MaLR"
repeat_region   35868. .35939
                /rpt_family="MIR"
repeat_region   38314. .38622
                /rpt_family="L1"
repeat_region   38628. .38765
                /rpt_family="L1"
repeat_region   38872. .39419
                /rpt_family="L1"
repeat_region   41647. .41946
                /rpt_family="Alu"
repeat_region   42051. .42251
                /rpt_family="L1"
repeat_region   42269. .42886
                /rpt_family="ERV1"
repeat_region   42887. .43631
                /rpt_family="L1"
repeat_region   43803. .44108
                /rpt_family="Alu"
repeat_region   47859. .47918
                /rpt_family="ERVL"
repeat_region   47963. .48092
                /rpt_family="ERVL"
repeat_region   48723. .49085
                /rpt_family="ERV1"
repeat_region   49086. .49402
                /rpt_family="Alu"
repeat_region   49403. .49437
                /rpt_family="ERV1"
repeat_region   50693. .51001
                /rpt_family="Alu"
repeat_region   51495. .51598

  Query Match             60.0%;  Score 22.8;  DB 87;  Length 196721;
  Best Local Similarity   79.4%;  Pred. No. 2.7e+02;
  Matches   27;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       5 ttttgctcaaagtttaaactttgagcaaaagcct 38
           | ||||||||||||  |  ||||| || ||||||
Db   44473 TGTTGCTCAAAGTTGTATGTTTGATCAGAAGCCT 44506


RESULT  18
AC019179/c
LOCUS       AC019179   196721 bp    DNA             PRI       08-NOV-2000
DEFINITION  Homo sapiens BAC clone RP11-240A16 from 4, complete sequence.
ACCESSION   AC019179
VERSION     AC019179.4  GI:11120947
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 196721)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 196721)
  AUTHORS   Harkins,R., Maupin,R., Gregory,S., Coblitz,B. and Fleming,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-240A16
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 196721)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 196721)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2000) Department of Genetics, Washington
```

```
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Nov 8, 2000 this sequence version replaced gi:7630907.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0240A16
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            Actual start of this clone is at base position 1 of RP11-240A16;
            actual end is at base position 196721 of RP11-240A16.

            The sequence H_NH0240A16 from base positon 157677 to 158503
            contains a tandem repeat.  The assembly is consistant with digest
            informationbut the sequence fidelity cannot be guaranteed.
FEATURES                    Location/Qualifiers
     source                 1. .196721
                            /organism="Homo sapiens"
                            /db_xref="taxon:9606"
                            /chromosome="4"
                            /map="4"
                            /clone="RP11-240A16"
                            /clone_lib="RPCI-11"
     repeat_region          3448. .3840
                            /rpt_family="ERV"
     repeat_region          4103. .4596
                            /rpt_family="ERVK"
     repeat_region          5164. .5231
                            /rpt_family="MIR"
     repeat_region          5616. .5734
                            /rpt_family="MIR"
     repeat_region          5861. .5911
                            /rpt_family="Mariner"
     repeat_region          6619. .6749
                            /rpt_family="MaLR"
     repeat_region          6727. .6776
                            /rpt_family="MaLR"
     repeat_region          8458. .8684
                            /rpt_family="Alu"
     repeat_region          8685. .8843
                            /rpt_family="Alu"
     repeat_region          10345. .10912
                            /rpt_family="MaLR"
     repeat_region          10952. .11592
                            /rpt_family="L2"
     repeat_region          14824. .15153
                            /rpt_family="L1"
     repeat_region          16195. .16619
                            /rpt_family="ERVL"
     repeat_region          16649. .16822
                            /rpt_family="MaLR"
     repeat_region          17787. .17850
                            /rpt_family="Alu"
     repeat_region          18581. .18773
                            /rpt_family="MER1_type"
     repeat_region          19637. .19697
                            /rpt_family="L2"
     repeat_region          19798. .19971
                            /rpt_family="Alu"
     repeat_region          19991. .20128
                            /rpt_family="Alu"
     repeat_region          22389. .22580
                            /rpt_family="MIR"
     repeat_region          23974. .24287
                            /rpt_family="Alu"
     repeat_region          25165. .25399
                            /rpt_family="MIR"
     repeat_region          25434. .25719
                            /rpt_family="Alu"
     repeat_region          26014. .26157
                            /rpt_family="L2"
     repeat_region          28448. .28553
                            /rpt_family="MIR"
     repeat_region          29093. .29243
                            /rpt_family="Alu"
     repeat_region          29715. .30240
                            /rpt_family="MaLR"
     repeat_region          30717. .30823
                            /rpt_family="L1"
     repeat_region          30859. .30929
                            /rpt_family="L1"
     repeat_region          30935. .31516
                            /rpt_family="ERV1"
     repeat_region          31502. .31764
                            /rpt_family="ERV1"
     repeat_region          31871. .31981
                            /rpt_family="L1"
     repeat_region          31982. .32494
                            /rpt_family="ERVK"
     repeat_region          32495. .32523
                            /rpt_family="L1"
     repeat_region          32579. .32646
                            /rpt_family="MaLR"
     repeat_region          32775. .32922
                            /rpt_family="ERVL"
     repeat_region          33108. .33153
                            /rpt_family="ERVL"
     repeat_region          33187. .33486
                            /rpt_family="Alu"
     repeat_region          33588. .33696
                            /rpt_family="MaLR"
     repeat_region          34154. .34461
                            /rpt_family="L2"
     repeat_region          34462. .35265
                            /rpt_family="ERV1"
     repeat_region          35266. .35432
                            /rpt_family="L2"
     repeat_region          35517. .35685
                            /rpt_family="MaLR"
     repeat_region          35868. .35939
                            /rpt_family="MIR"
     repeat_region          38314. .38622
                            /rpt_family="L1"
     repeat_region          38628. .38765
```

```
                     /rpt_family="L1"
     repeat_region   38872. .39419
                     /rpt_family="L1"
     repeat_region   41647. .41946
                     /rpt_family="Alu"
     repeat_region   42051. .42251
                     /rpt_family="L1"
     repeat_region   42269. .42886
                     /rpt_family="ERV1"
     repeat_region   42887. .43631
                     /rpt_family="L1"
     repeat_region   43803. .44108
                     /rpt_family="Alu"
     repeat_region   47859. .47918
                     /rpt_family="ERVL"
     repeat_region   47963. .48092
                     /rpt_family="ERVL"
     repeat_region   48723. .49085
                     /rpt_family="ERV1"
     repeat_region   49086. .49402
                     /rpt_family="Alu"
     repeat_region   49403. .49437
                     /rpt_family="ERV1"
     repeat_region   50693. .51001
                     /rpt_family="Alu"
     repeat_region   51495. .51598

  Query Match             60.0%;  Score 22.8;  DB 87;  Length 196721;
  Best Local Similarity   79.4%;  Pred. No. 2.7e+02;
  Matches   27;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 aggcttttgctcaaagtttaaactttgagcaaaa 34
           |||||| || |||||  |  ||||||||||||
Db   44506 AGGCTTCTGATCAAACATACAACTTTGAGCAACA 44473


RESULT  19
SSU70240
LOCUS       SSU70240     4267 bp    DNA             MAM       10-AUG-1998
DEFINITION  Sus scrofa apolipoprotein-E (Apo-E) gene, complete cds.
ACCESSION   U70240
VERSION     U70240.1  GI:2388608
KEYWORDS    .
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 4267)
  AUTHORS   Ramsoondar,J.J., Rucker,E.B., Vasquez,J.C., Gallagher,D.S.,
            Grimm,D.R., Lunney,J.K., Schook,L.B. and Piedrahita,J.A.
  TITLE     Isolation and genetic characterization of the porcine
            apolipoprotein E gene
  JOURNAL   Anim. Genet. 29 (1), 43-47 (1998)
  MEDLINE   98347372
REFERENCE   2  (bases 1 to 4267)
  AUTHORS   Ramsoondar,J.J. and Piedrahita,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1996) VAPH, Texas A&M University, College
            Station, TX 77843, USA
FEATURES             Location/Qualifiers
     source          1. .4267
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
     exon            832. .857
                     /gene="Apo-E"
                     /number=1
     gene            832. .3879
                     /gene="Apo-E"
     mRNA            join(832. .857,1686. .1728,2473. .2662,3037. .3879)
                     /gene="Apo-E"
     intron          858. .1662
                     /gene="Apo-E"
                     /number=1
     exon            1663. .1728
                     /gene="Apo-E"
                     /number=2
     CDS             join(1686. .1728,2473. .2662,3037. .3757)
                     /gene="Apo-E"
                     /note="plasma lipoprotein"
                     /codon_start=1
                     /product="apolipoprotein-E"
                     /protein_id="AAC29512.1"
                     /db_xref="GI:2388609"
                     /translation="MRVLWVALVVTLLAGCRTEDEPGPPPEVHVWWEEPKWQGSQPWE
                     QALGRFWDYLRWVQSLSDQVQEELLSTKVTQELTELIEESMKEVKAYREELEAQLGPV
                     TQETQARLSKELQAAQARVGADMEDVRNRLVLYRSEVHNMLGQTTEELRSRLASHLRK
                     LRKRLLRDTEDLQKRLAVYQAGLREGAERSVSALRERLGPLVEQGRLRAATLSTRAGQ
                     PLRERAEAWGQKLRGRLEEMGSRTRDRLDEMREQLEEVRTKVEEQGSQLRLQAEGFHA
                     LLKGWFEPLVEDIRRQWAGLVERMQSGVSISSSTSAPSDNQ"
     intron          1729. .2472
                     /gene="Apo-E"
                     /number=2
     exon            2473. .2662
                     /gene="Apo-E"
                     /number=3
     intron          2663. .3036
                     /gene="Apo-E"
                     /number=3
     exon            3037. .3879
                     /gene="Apo-E"
                     /number=4
BASE COUNT      841 a   1236 c   1344 g    846 t
ORIGIN


  Query Match             59.5%;  Score 22.6;  DB 7;  Length 4267;
  Best Local Similarity   86.2%;  Pred. No. 6.9e+02;
  Matches   25;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      10 ctcaaagtttaaactttgagcaaaagcct 38
           |||||||  |||||||||||||| | ||||
Db     332 CTCAAAGCCTAAACTTTGAGCAGATGCCT 360


RESULT  20
SSU70240/c
LOCUS       SSU70240     4267 bp    DNA             MAM       10-AUG-1998
DEFINITION  Sus scrofa apolipoprotein-E (Apo-E) gene, complete cds.
ACCESSION   U70240
VERSION     U70240.1  GI:2388608
KEYWORDS    .
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 4267)
  AUTHORS   Ramsoondar,J.J., Rucker,E.B., Vasquez,J.C., Gallagher,D.S.,
            Grimm,D.R., Lunney,J.K., Schook,L.B. and Piedrahita,J.A.
  TITLE     Isolation and genetic characterization of the porcine
            apolipoprotein E gene
  JOURNAL   Anim. Genet. 29 (1), 43-47 (1998)
  MEDLINE   98347372
REFERENCE   2  (bases 1 to 4267)
  AUTHORS   Ramsoondar,J.J. and Piedrahita,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1996) VAPH, Texas A&M University, College
            Station, TX 77843, USA
FEATURES             Location/Qualifiers
     source          1. .4267
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
     exon            832. .857
                     /gene="Apo-E"
                     /number=1
     gene            832. .3879
                     /gene="Apo-E"
```

```
     mRNA            join(832. .857,1686. .1728,2473. .2662,3037. .3879)
                     /gene="Apo-E"
     intron          858. .1662
                     /gene="Apo-E"
                     /number=1
     exon            1663. .1728
                     /gene="Apo-E"
                     /number=2
     CDS             join(1686. .1728,2473. .2662,3037. .3757)
                     /gene="Apo-E"
                     /note="plasma lipoprotein"
                     /codon_start=1
                     /product="apolipoprotein-E"
                     /protein_id="AAC29512.1"
                     /db_xref="GI:2388609"
                     /translation="MRVLWVALVVTLLAGCRTEDEPGPPPEVHVWWEEPKWQGSQPWE
                     QALGRFWDYLRWVQSLSDQVQEELLSTKVTQELTELIEESMKEVKAYREELEAQLGPV
                     TQETQARLSKELQAAQARVGADMEDVRNRLVLYRSEVHNMLGQTTEELRSRLASHLRK
                     LRKRLLRDTEDLQKRLAVYQAGLREGAERSVSALRERLGPLVEQGRLRAATLSTRAGQ
                     PLRERAEAWGQKLRGRLEEMGSRTRDRLDEMREQLEEVRTKVEEQGSQLRLQAEGFHA
                     LLKGWFEPLVEDIRRQWAGLVERMQSGVSISSSTSAPSDNQ"
     intron          1729. .2472
                     /gene="Apo-E"
                     /number=2
     exon            2473. .2662
                     /gene="Apo-E"
                     /number=3
     intron          2663. .3036
                     /gene="Apo-E"
                     /number=3
     exon            3037. .3879
                     /gene="Apo-E"
                     /number=4
BASE COUNT      841 a   1236 c   1344 g    846 t
ORIGIN


  Query Match           59.5%;  Score 22.6;  DB 7;  Length 4267;
  Best Local Similarity  86.2%;  Pred. No. 6.9e+02;
  Matches   25;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgag 29
          |||| | ||||||||||||  |||||||
Db    360 AGGCATCTGCTCAAAGTTTAGGCTTTGAG 332


Search completed: June  7, 2001, 18:06:14
Job time: 6367 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 7, 2001, 18:09:57 ; Search time 200.8 Seconds
(without alignments)
110.477 Million cell updates/sec

Title: US-09-601-561-3
Perfect score: 38
Sequence: 1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : N_Geneseq_0401:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	21.8	57.4	1256	21	C49875	Arabidopsis thalia
c	2	21.8	57.4	1256	21	C49875	Arabidopsis thalia
	3	21.8	57.4	1257	21	C47834	Arabidopsis thalia
c	4	21.8	57.4	1257	21	C47834	Arabidopsis thalia
	5	20.8	54.7	1319	21	A38558	Actinobacillus ple
c	6	20.8	54.7	1319	21	A38558	Actinobacillus ple
	7	20.8	54.7	2000	19	V33669	Candida glabrata I
c	8	20.8	54.7	2000	19	V33669	Candida glabrata I
	9	20.6	54.2	252	21	C31525	Human secreted pro
c	10	20.6	54.2	252	21	C31525	Human secreted pro
	11	20.6	54.2	714	21	A68248	Bacteriophage 77 7
c	12	20.6	54.2	714	21	A68248	Bacteriophage 77 7
	13	20.6	54.2	1374	21	F15923	Human prostate can
c	14	20.6	54.2	1374	21	F15923	Human prostate can
	15	20.6	54.2	41708	21	A68247	Bacteriophage 77 c
c	16	20.6	54.2	41708	21	A68247	Bacteriophage 77 c
c	17	20.4	53.7	1664976	19	V21209	Methanococcus jann
	18	20	52.6	1593	20	X06832	Disease associated
c	19	20	52.6	1593	20	X06832	Disease associated
	20	20	52.6	1833	19	V46339	Serine threonine k
c	21	20	52.6	1833	19	V46339	Serine threonine k
	22	20	52.6	20561	20	X13229	Enterococcus faeca
c	23	20	52.6	20561	20	X13229	Enterococcus faeca
	24	19.8	52.1	1353	21	Z51685	C. elegans alpha-1
c	25	19.8	52.1	1353	21	Z51685	C. elegans alpha-1
	26	19.6	51.6	267	18	T67414	H. pylori transpor
c	27	19.6	51.6	267	18	T67414	H. pylori transpor
	28	19.6	51.6	669	19	X30436	H. pylori cellular
c	29	19.6	51.6	669	19	X30436	H. pylori cellular
	30	19.6	51.6	708	18	T68277	H. pylori transpor
c	31	19.6	51.6	708	18	T68277	H. pylori transpor
	32	19.6	51.6	6365	14	Q41173	Plasmid LXSN-RI-IL
c	33	19.6	51.6	6365	14	Q41173	Plasmid LXSN-RI-IL
	34	19.2	50.5	1110	21	C79664	Virulence gene #71
c	35	19.2	50.5	1110	21	C79664	Virulence gene #71
	36	19.2	50.5	6145	20	X84569	Biorhythm marker g
c	37	19.2	50.5	6145	20	X84569	Biorhythm marker g
	38	19	50.0	255	20	V89341	EST clone CJ206.
c	39	19	50.0	255	20	V89341	EST clone CJ206.
	40	19	50.0	1312	21	C40096	Arabidopsis thalia
c	41	19	50.0	1312	21	C40096	Arabidopsis thalia
	42	19	50.0	2006	21	A26363	Human secreted pro
c	43	19	50.0	2006	21	A26363	Human secreted pro
	44	19	50.0	7958	21	Z34835	Mouse integrin sub
c	45	19	50.0	7958	21	Z34835	Mouse integrin sub
	46	18.8	49.5	350	20	V88989	EST clone IJ597.
c	47	18.8	49.5	350	20	V88989	EST clone IJ597.
	48	18.8	49.5	534	18	T68264	H. pylori cell env
c	49	18.8	49.5	534	18	T68264	H. pylori cell env
	50	18.8	49.5	560	21	C41961	Arabidopsis thalia

ALIGNMENTS

RESULT 1
C49875
ID C49875 standard; DNA; 1256 BP.
XX
AC C49875;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 62768.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.

PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 57.4%; Score 21.8; DB 21; Length 1256;
Best Local Similarity 78.8%; Pred. No. 20;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy      3 gcttttgctcaaagtttaaactttgagcaaaag 35
          |||||| |||||  || ||||||||  | ||||
Db   1179 gctttttctcaagtttaaaactttgttctaaag 1211
```

RESULT 2
C49875/c
ID C49875 standard; DNA; 1256 BP.
XX
AC C49875;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 62768.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 57.4%; Score 21.8; DB 21; Length 1256;
Best Local Similarity 78.8%; Pred. No. 20;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy      4 cttttgctcaaagtttaaactttgagcaaaagc 36
          |||| |  |||||||| ||  ||||| ||||||
Db   1211 CTTTAGAACAAAGTTTTAAACTTGAGAAAAAGC 1179
```

RESULT 3
C47834
ID C47834 standard; DNA; 1257 BP.
XX
AC C47834;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 55292.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 57.4%; Score 21.8; DB 21; Length 1257;
Best Local Similarity 78.8%; Pred. No. 20;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy       3 gcttttgctcaaagtttaaactttgagcaaaag 35
           |||||| |||||  || ||||||||  | ||||
Db    1177 gctttttctcaagtttaaaactttgttctaaag 1209
```

RESULT 4
C47834/c
ID C47834 standard; DNA; 1257 BP.
XX
AC C47834;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 55292.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 57.4%; Score 21.8; DB 21; Length 1257;
Best Local Similarity 78.8%; Pred. No. 20;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy      4 cttttgctcaaagtttaaactttgagcaaaagc 36
          |||| |  |||||||| ||  ||||| ||||||
Db   1209 CTTTAGAACAAAGTTTTAAACTTGAGAAAAAGC 1177
```

RESULT 5
A38558
ID A38558 standard; DNA; 1319 BP.
XX
AC A38558;
XX
DT 11-SEP-2000 (first entry)
XX
DE Actinobacillus pleuropneumoniae ompA2 DNA.
XX
KW OmpA2; outer membrane protein; APP; pneumonic pathogen; swine;
KW Sus scrofa; serotype; antigen; vaccine; cross-reactive immunity;
KW cross-protection; diagnosis; ds.
XX
OS Actinobacillus pleuropneumoniae strain Pz420.
XX
FH Key Location/Qualifiers
FT CDS 197..1306
FT /*tag= a
FT /product= "Actinobacillus pleuropneumoniae (APP) OmpA2"
FT sig_peptide 297..253
FT /*tag= b
FT mat_peptide 254..1303
FT /*tag= c
XX
PN EP1001025-A2.
XX
PD 17-MAY-2000.
XX
PF 20-OCT-1999; 99EP-0308262.
XX
PR 22-OCT-1998; 98US-0105285.
XX
PA (PFIZ) PFIZER PROD INC.
XX
PI Ankenbauer RG, Baarsch MJ, Campos M, Keich RL, Rosey EL;
PI Warren-Stewart LM, Suiter BT;
XX
DR WPI; 2000-320438/28.
DR P-PSDB; Y97900.
XX
PT Low molecular weight Actinobacillus pleuropneumoniae proteins and DNA
PT encoding them, for use as vaccines against the bacteria in swine -
XX
PS Claim 32; Page 49-51; 81pp; English.
XX
CC This sequence represents DNA encoding a low molecular weight outer
CC membrane protein, OmpA2, from Actinobacillus pleuropneumoniae (APP)
CC strain Pz420 (ATCC 98930). The invention relates to the novel APP outer
CC membrane proteins Omp20, OmpW, Omp27, OmpA1 and OmpA2 (Y97896-Y97900)
CC and to nucleic acids encoding them (A38554-A38558). APP is a Gram
CC negative coccobacillus which is one of the most important swine pneumonic
CC pathogens. 12 different serotypes of APP have been recognised which vary
CC in geographic distribution. Prior art attempts at vaccinating against
CC APP have produced mainly serotype-specific immune responses. In contrast,
CC natural immunity to any one serotype seems to confer significant
CC protection from disease caused by other serotypes, suggesting that

```
CC   natural exposure induces cross-reactive immunity to shared antigens. The
CC   novel outer membrane proteins of the invention are present in all 12
CC   serotypes, and may provide a target for cross-protective immunisation.
CC   The novel outer membrane proteins and nucleic acids encoding them can be
CC   used as a vaccine against APP in swine. They can also be used as reagents
CC   for the diagnosis of APP infections.
XX
SQ   Sequence 1319 BP; 428 A; 252 C; 260 G; 379 T; 0 other;


  Query Match             54.7%;  Score 20.8;  DB 21;  Length 1319;
  Best Local Similarity   78.1%;  Pred. No. 48;
  Matches   25;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      6 tttgctcaaagtttaaactttgagcaaaagcc 37
          | ||||||  |||||||||||| |  ||| ||
Db    521 tctgctcacggtttaaactttgcgttaaaacc 552


RESULT   6
A38558/c
ID   A38558 standard; DNA; 1319 BP.
XX
AC   A38558;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Actinobacillus pleuropneumoniae ompA2 DNA.
XX
KW   OmpA2; outer membrane protein; APP; pneumonic pathogen; swine;
KW   Sus scrofa; serotype; antigen; vaccine; cross-reactive immunity;
KW   cross-protection; diagnosis; ds.
XX
OS   Actinobacillus pleuropneumoniae strain Pz420.
XX
FH   Key             Location/Qualifiers
FT   CDS             197..1306
FT                   /*tag=  a
FT                   /product= "Actinobacillus pleuropneumoniae (APP) OmpA2"
FT   sig_peptide     297..253
FT                   /*tag=  b
FT   mat_peptide     254..1303
FT                   /*tag=  c
XX
PN   EP1001025-A2.
XX
PD   17-MAY-2000.
XX
PF   20-OCT-1999;   99EP-0308262.
XX
PR   22-OCT-1998;   98US-0105285.
XX
PA   (PFIZ ) PFIZER PROD INC.
XX
PI   Ankenbauer RG,  Baarsch MJ,  Campos M,  Keich RL,  Rosey EL;
PI   Warren-Stewart LM,  Suiter BT;
XX
DR   WPI; 2000-320438/28.
DR   P-PSDB; Y97900.
XX
PT   Low molecular weight Actinobacillus pleuropneumoniae proteins and DNA
PT   encoding them, for use as vaccines against the bacteria in swine -
XX
PS   Claim 32; Page 49-51; 81pp; English.
XX
CC   This sequence represents DNA encoding a low molecular weight outer
CC   membrane protein, OmpA2, from Actinobacillus pleuropneumoniae (APP)
CC   strain Pz420 (ATCC 98930). The invention relates to the novel APP outer
CC   membrane proteins Omp20, OmpW, Omp27, OmpA1 and OmpA2 (Y97896-Y97900)
CC   and to nucleic acids encoding them (A38554-A38558). APP is a Gram
CC   negative coccobacillus which is one of the most important swine pneumonic
CC   pathogens. 12 different serotypes of APP have been recognised which vary
CC   in geographic distribution. Prior art attempts at vaccinating against
CC   APP have produced mainly serotype-specific immune responses. In contrast,
CC   natural immunity to any one serotype seems to confer significant
CC   protection from disease caused by other serotypes, suggesting that
CC   natural exposure induces cross-reactive immunity to shared antigens. The
CC   novel outer membrane proteins of the invention are present in all 12
CC   serotypes, and may provide a target for cross-protective immunisation.
CC   The novel outer membrane proteins and nucleic acids encoding them can be
CC   used as a vaccine against APP in swine. They can also be used as reagents
CC   for the diagnosis of APP infections.
XX
SQ   Sequence 1319 BP; 428 A; 252 C; 260 G; 379 T; 0 other;


  Query Match             54.7%;  Score 20.8;  DB 21;  Length 1319;
  Best Local Similarity   78.1%;  Pred. No. 48;
  Matches   25;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      2 ggcttttgctcaaagtttaaactttgagcaaa 33
            || |||  | ||||||||||||  |||||| |
Db    552 GGTTTTAACGCAAAGTTTAAACCGTGAGCAGA 521


RESULT   7
V33669
ID   V33669 standard; DNA; 2000 BP.
XX
AC   V33669;
XX
DT   19-JAN-1999  (first entry)
XX
DE   Candida glabrata IPC synthase encoding DNA.
XX
KW   Candida; IPC synthase; fungus; inositolphosphoryl ceramide synthase;
KW   anti-fungal therapy; sphingolipid biosynthesis; phosphatidylinositol; ss.
XX
OS   Candida glabrata.
XX
FH   Key             Location/Qualifiers
FT   CDS             501..1751
FT                   /*tag=  a
XX
PN   EP872485-A2.
XX
PD   21-OCT-1998.
XX
PF   14-APR-1998;   98EP-0302866.
XX
PR   17-OCT-1997;   97US-0062971.
PR   15-APR-1997;   97US-0043591.
PR   22-APR-1997;   97US-0044095.
PR   13-MAY-1997;   97US-0046348.
PR   21-JUL-1997;   97US-0053320.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Heider SA,  Radding JA;
XX
DR   WPI; 1998-533879/46.
DR   P-PSDB; W70514.
XX
PT   New inositolphosphoryl ceramide synthase genes from fungi - useful
PT   for identifying compounds for anti-fungal therapy
XX
PS   Claim 5; Page 10-12; 53pp; English.
XX
CC   The present sequence encodes a pure inositolphosphoryl ceramide (IPC)
CC   synthase protein from a fungal cell, Candida glabrata. The present
CC   invention also describes a method for identifying inhibitory compounds
CC   of fungal IPC synthase protein activity. IPC synthase proteins are
CC   useful for identifying inhibitors of fungal sphingolipid biosynthesis,
CC   as the IPC synthase catalyses a step in the synthesis of
CC   inositolphosphoryl ceramide from ceramide and phosphatidylinositol.
CC   Fragments of IPC synthase proteins are also useful as probes or primers
```

```
CC   for identification and isolation of homologous genes.
XX
SQ   Sequence 2000 BP; 609 A; 365 C; 324 G; 702 T; 0 other;


  Query Match             54.7%;  Score 20.8;  DB 19;  Length 2000;
  Best Local Similarity   78.1%;  Pred. No. 50;
  Matches   25;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       5 ttttgctcaaagtttaaactttgagcaaaagc 36
           |||| |||||| |||   ||||||| |||| |
Db    1918 tttttctcaaaatttgccctttgagaaaaacc 1949


RESULT   8
V33669/c
ID   V33669 standard; DNA; 2000 BP.
XX
AC   V33669;
XX
DT   19-JAN-1999  (first entry)
XX
DE   Candida glabrata IPC synthase encoding DNA.
XX
KW   Candida; IPC synthase; fungus; inositolphosphoryl ceramide synthase;
KW   anti-fungal therapy; sphingolipid biosynthesis; phosphatidylinositol; ss.
XX
OS   Candida glabrata.
XX
FH   Key             Location/Qualifiers
FT   CDS             501..1751
FT                   /*tag=  a
XX
PN   EP872485-A2.
XX
PD   21-OCT-1998.
XX
PF   14-APR-1998;    98EP-0302866.
XX
PR   17-OCT-1997;    97US-0062971.
PR   15-APR-1997;    97US-0043591.
PR   22-APR-1997;    97US-0044095.
PR   13-MAY-1997;    97US-0046348.
PR   21-JUL-1997;    97US-0053320.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Heider SA,  Radding JA;
XX
DR   WPI; 1998-533879/46.
DR   P-PSDB; W70514.
XX
PT   New inositolphosphoryl ceramide synthase genes from fungi - useful
PT   for identifying compounds for anti-fungal therapy
XX
PS   Claim 5; Page 10-12; 53pp; English.
XX
CC   The present sequence encodes a pure inositolphosphoryl ceramide (IPC)
CC   synthase protein from a fungal cell, Candida glabrata. The present
CC   invention also describes a method for identifying inhibitory compounds
CC   of fungal IPC synthase protein activity. IPC synthase proteins are
CC   useful for identifying inhibitors of fungal sphingolipid biosynthesis,
CC   as the IPC synthase catalyses a step in the synthesis of
CC   inositolphosphoryl ceramide from ceramide and phosphatidylinositol.
CC   Fragments of IPC synthase proteins are also useful as probes or primers
CC   for identification and isolation of homologous genes.
XX
SQ   Sequence 2000 BP; 609 A; 365 C; 324 G; 702 T; 0 other;


  Query Match             54.7%;  Score 20.8;  DB 19;  Length 2000;
  Best Local Similarity   78.1%;  Pred. No. 50;
  Matches   25;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3 gcttttgctcaaagtttaaactttgagcaaaa 34
           | |||| |||||||   ||| |||||| ||||
Db    1949 GGTTTTTCTCAAAGGGCAAATTTTGAGAAAAA 1918


RESULT   9
C31525
ID   C31525 standard; cDNA; 252 BP.
XX
AC   C31525;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 35600.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 35600; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 252 BP; 89 A; 35 C; 35 G; 92 T; 1 other;


  Query Match             54.2%;  Score 20.6;  DB 21;  Length 252;
  Best Local Similarity   74.3%;  Pred. No. 50;
  Matches   26;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 aggcttttgctcaaagtttaaactttgagcaaaag 35
             || ||| ||  | | |||||||| ||||  ||||||
Db     152 agacttatgaaccacgtttaaaatttgtacaaaag 186


RESULT  10
C31525/c
ID   C31525 standard; cDNA; 252 BP.
XX
AC   C31525;
```

XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 35600.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 35600; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 252 BP; 89 A; 35 C; 35 G; 92 T; 1 other;

Query Match 54.2%; Score 20.6; DB 21; Length 252;
Best Local Similarity 74.3%; Pred. No. 50;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 cttttgctcaaagtttaaactttgagcaaaagcct 38
|||||| |||| ||||||| | | || ||| ||
Db 186 CTTTTGTACAAATTTTAAACGTGGTTCATAAGTCT 152

RESULT 11
A68248
ID A68248 standard; DNA; 714 BP.
XX
AC A68248;
XX
DT 27-OCT-2000 (first entry)
XX
DE Bacteriophage 77 77ORF017 nucleotide sequence.
XX
KW Bacteriophage; antimicrobial; genome; identification; antibacterial;
KW bacterial growth inhibition; bacterial infection; ds.
XX
OS Bacteriophage 77.
XX
PN WO200032825-A2.
XX
PD 08-JUN-2000.
XX
PF 03-DEC-1999; 99WO-IB02040.
XX
PR 03-DEC-1998; 98US-0110992.
PR 03-JUN-1999; 99US-0326144.
PR 28-SEP-1999; 99US-0407804.
PR 30-SEP-1999; 99US-0157218.
PR 01-DEC-1999; 99US-0168777.
PR 02-DEC-1999; 99US-0454252.
XX
PA (PHAG-) PHAGETECH INC.
XX
PI Pelletier J, Gros P, Dubow M;
XX
DR WPI; 2000-412361/35.
DR P-PSDB; B16523.
XX
PT Identifying a bacteriophage coding region for treating bacterial
PT infections comprises identifying a nucleic acid encoding a product that
PT inhibits bacteria when a bacteriophage infects a bacterium -
XX
PS Disclosure; Page 153; 456pp; English.
XX
CC The present invention describes a method for identifying a bacteriophage
CC coding region encoding a product active on an essential bacterial target.
CC The method comprises identifying a nucleic acid sequence encoding a gene
CC product that provides a bacteria-inhibiting function when an
CC uncharacterised bacteriophage infects a pathogenic bacterium. The
CC compound active on a target of a bacteriophage inhibitor protein in a
CC bacteria is used to treat or prevent a bacterial infection in an animal.
CC A68243 to A69442 and B16523 to B16954 represent bacteriophage nucleotide
CC and protein sequences which are used in the exemplification of the
CC present invention.
XX
SQ Sequence 714 BP; 312 A; 77 C; 96 G; 229 T; 0 other;

Query Match 54.2%; Score 20.6; DB 21; Length 714;
Best Local Similarity 74.3%; Pred. No. 54;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 aggcttttgctcaaagtttaaactttgagcaaaag 35
||| |||| |||||||||||| ||| || ||
Db 123 aggtttttatccaaagtttaaacaaggagaaatag 157

RESULT 12
A68248/c
ID A68248 standard; DNA; 714 BP.
XX
AC A68248;
XX
DT 27-OCT-2000 (first entry)
XX
DE Bacteriophage 77 77ORF017 nucleotide sequence.
XX
KW Bacteriophage; antimicrobial; genome; identification; antibacterial;
KW bacterial growth inhibition; bacterial infection; ds.
XX
OS Bacteriophage 77.
XX
PN WO200032825-A2.
XX
PD 08-JUN-2000.
XX
PF 03-DEC-1999; 99WO-IB02040.
XX
PR 03-DEC-1998; 98US-0110992.
PR 03-JUN-1999; 99US-0326144.
PR 28-SEP-1999; 99US-0407804.
PR 30-SEP-1999; 99US-0157218.

PR 01-DEC-1999; 99US-0168777.
PR 02-DEC-1999; 99US-0454252.
XX
PA (PHAG-) PHAGETECH INC.
XX
PI Pelletier J, Gros P, Dubow M;
XX
DR WPI; 2000-412361/35.
DR P-PSDB; B16523.
XX
PT Identifying a bacteriophage coding region for treating bacterial
PT infections comprises identifying a nucleic acid encoding a product that
PT inhibits bacteria when a bacteriophage infects a bacterium -
XX
PS Disclosure; Page 153; 456pp; English.
XX
CC The present invention describes a method for identifying a bacteriophage
CC coding region encoding a product active on an essential bacterial target.
CC The method comprises identifying a nucleic acid sequence encoding a gene
CC product that provides a bacteria-inhibiting function when an
CC uncharacterised bacteriophage infects a pathogenic bacterium. The
CC compound active on a target of a bacteriophage inhibitor protein in a
CC bacteria is used to treat or prevent a bacterial infection in an animal.
CC A68243 to A69442 and B16523 to B16954 represent bacteriophage nucleotide
CC and protein sequences which are used in the exemplification of the
CC present invention.
XX
SQ Sequence 714 BP; 312 A; 77 C; 96 G; 229 T; 0 other;

Query Match 54.2%; Score 20.6; DB 21; Length 714;
Best Local Similarity 74.3%; Pred. No. 54;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 cttttgctcaaagtttaaactttgagcaaaagcct 38
|| || ||| |||||||||||| |||| |||
Db 157 CTATTTCTCCTTGTTTAAACTTTGGATAAAAACCT 123

RESULT 13
F15923
ID F15923 standard; cDNA; 1374 BP.
XX
AC F15923;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:358.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW vulnerary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease; ss.
XX
OS Homo sapiens.
XX
PN WO200055174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05988.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587513/55.
DR P-PSDB; B56720.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
PS Claim 1; Page 856-857; 2338pp; English.
XX
CC F15566 to F16505 encode the human prostate cancer associated proteins,
CC called prostate cancer antigens, given in B56363 to B57302. The prostate
CC cancer antigens can have neuroprotective, cytostatic, cardioactive,
CC immunomodulatory, muscular, vulnerary, gastrointestinal, nephrotropic,
CC antiinfective, gynaecological and antibacterial activities, and can be
CC used in gene therapy. The prostate cancer antigen polynucleotides may be
CC used for detection of prostate cancer, chromosome identification, as
CC chromosome markers, and for numerous other diagnostic or research
CC purposes. The prostate cancer antigens may be used to treat disorders
CC such as neural, immune, muscular, reproductive, gastrointestinal,
CC pulmonary, cardiovascular, renal, and proliferative disorders, wounds,
CC and infectious diseases. F16506 to F16514 to B57303 represent sequences
CC used in the exemplification of the present invention.
XX
SQ Sequence 1374 BP; 391 A; 268 C; 281 G; 432 T; 2 other;

Query Match 54.2%; Score 20.6; DB 21; Length 1374;
Best Local Similarity 74.3%; Pred. No. 57;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 aggcttttgctcaaagtttaaactttgagcaaaag 35
| | || | ||| | ||||| ||||||| |||||
Db 680 aagattattttcagaatttaatctttgaggaaaag 714

RESULT 14
F15923/c
ID F15923 standard; cDNA; 1374 BP.
XX
AC F15923;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:358.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW vulnerary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease; ss.
XX
OS Homo sapiens.
XX
PN WO200055174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05988.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587513/55.
DR P-PSDB; B56720.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
PS Claim 1; Page 856-857; 2338pp; English.

```
XX
CC   F15566 to F16505 encode the human prostate cancer associated proteins,
CC   called prostate cancer antigens, given in B56363 to B57302. The prostate
CC   cancer antigens can have neuroprotective, cytostatic, cardioactive,
CC   immunomodulatory, muscular, vulnerary, gastrointestinal, nephrotropic,
CC   antiinfective, gynaecological and antibacterial activities, and can be
CC   used in gene therapy. The prostate cancer antigen polynucleotides may be
CC   used for detection of prostate cancer, chromosome identification, as
CC   chromosome markers, and for numerous other diagnostic or research
CC   purposes. The prostate cancer antigens may be used to treat disorders
CC   such as neural, immune, muscular, reproductive, gastrointestinal,
CC   pulmonary, cardiovascular, renal, and proliferative disorders, wounds,
CC   and infectious diseases. F16506 to F16514 to B57303 represent sequences
CC   used in the exemplification of the present invention.
XX
SQ   Sequence 1374 BP; 391 A; 268 C; 281 G; 432 T; 2 other;


  Query Match             54.2%;  Score 20.6;  DB 21;  Length 1374;
  Best Local Similarity   74.3%;  Pred. No. 57;
  Matches   26;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      4 cttttgctcaaagtttaaactttgagcaaaagcct 38
          ||||| ||||||| ||||| | |||  | || | |
Db    714 CTTTTCCTCAAAGATTAAATTCTGAAAATAATCTT 680


RESULT  15
A68247
ID   A68247 standard; DNA; 41708 BP.
XX
AC   A68247;
XX
DT   27-OCT-2000  (first entry)
XX
DE   Bacteriophage 77 complete genome sequence.
XX
KW   Bacteriophage; antimicrobial; genome; identification; antibacterial;
KW   bacterial growth inhibition; bacterial infection; ds.
XX
OS   Bacteriophage 77.
XX
PN   WO200032825-A2.
XX
PD   08-JUN-2000.
XX
PF   03-DEC-1999;   99WO-IB02040.
XX
PR   03-DEC-1998;   98US-0110992.
PR   03-JUN-1999;   99US-0326144.
PR   28-SEP-1999;   99US-0407804.
PR   30-SEP-1999;   99US-0157218.
PR   01-DEC-1999;   99US-0168777.
PR   02-DEC-1999;   99US-0454252.
XX
PA   (PHAG-) PHAGETECH INC.
XX
PI   Pelletier J,  Gros P,  Dubow M;
XX
DR   WPI; 2000-412361/35.
XX
PT   Identifying a bacteriophage coding region for treating bacterial
PT   infections comprises identifying a nucleic acid encoding a product that
PT   inhibits bacteria when a bacteriophage infects a bacterium     -
XX
PS   Example 3; Page 141-151; 456pp; English.
XX
CC   The present invention describes a method for identifying a bacteriophage
CC   coding region encoding a product active on an essential bacterial target.
CC   The method comprises identifying a nucleic acid sequence encoding a gene
CC   product that provides a bacteria-inhibiting function when an
CC   uncharacterised bacteriophage infects a pathogenic bacterium. The
CC   compound active on a target of a bacteriophage inhibitor protein in a
CC   bacteria is used to treat or prevent a bacterial infection in an animal.
CC   A68243 to A69442 and B16523 to B16954 represent bacteriophage nucleotide
CC   and protein sequences which are used in the exemplification of the
CC   present invention.
XX
SQ   Sequence 41708 BP; 15607 A; 5898 C; 8088 G; 12115 T; 0 other;


  Query Match             54.2%;  Score 20.6;  DB 21;  Length 41708;
  Best Local Similarity   74.3%;  Pred. No. 75;
  Matches   26;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      4 cttttgctcaaagtttaaactttgagcaaaagcct 38
          || || |||   ||||||||||||   |||| |||
Db  23826 ctatttctccttgtttaaactttggataaaaacct 23860


RESULT  16
A68247/c
ID   A68247 standard; DNA; 41708 BP.
XX
AC   A68247;
XX
DT   27-OCT-2000  (first entry)
XX
DE   Bacteriophage 77 complete genome sequence.
XX
KW   Bacteriophage; antimicrobial; genome; identification; antibacterial;
KW   bacterial growth inhibition; bacterial infection; ds.
XX
OS   Bacteriophage 77.
XX
PN   WO200032825-A2.
XX
PD   08-JUN-2000.
XX
PF   03-DEC-1999;   99WO-IB02040.
XX
PR   03-DEC-1998;   98US-0110992.
PR   03-JUN-1999;   99US-0326144.
PR   28-SEP-1999;   99US-0407804.
PR   30-SEP-1999;   99US-0157218.
PR   01-DEC-1999;   99US-0168777.
PR   02-DEC-1999;   99US-0454252.
XX
PA   (PHAG-) PHAGETECH INC.
XX
PI   Pelletier J,  Gros P,  Dubow M;
XX
DR   WPI; 2000-412361/35.
XX
PT   Identifying a bacteriophage coding region for treating bacterial
PT   infections comprises identifying a nucleic acid encoding a product that
PT   inhibits bacteria when a bacteriophage infects a bacterium     -
XX
PS   Example 3; Page 141-151; 456pp; English.
XX
CC   The present invention describes a method for identifying a bacteriophage
CC   coding region encoding a product active on an essential bacterial target.
CC   The method comprises identifying a nucleic acid sequence encoding a gene
CC   product that provides a bacteria-inhibiting function when an
CC   uncharacterised bacteriophage infects a pathogenic bacterium. The
CC   compound active on a target of a bacteriophage inhibitor protein in a
CC   bacteria is used to treat or prevent a bacterial infection in an animal.
CC   A68243 to A69442 and B16523 to B16954 represent bacteriophage nucleotide
CC   and protein sequences which are used in the exemplification of the
CC   present invention.
XX
SQ   Sequence 41708 BP; 15607 A; 5898 C; 8088 G; 12115 T; 0 other;


  Query Match             54.2%;  Score 20.6;  DB 21;  Length 41708;
  Best Local Similarity   74.3%;  Pred. No. 75;
```

Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Qy      1 aggcttttgctcaaagtttaaactttgagcaaaag 35
          ||| ||||   ||||||||||||   ||| || ||
Db  23860 AGGTTTTTATCCAAAGTTTAAACAAGGAGAAATAG 23826
```

RESULT 17
V21209/c
ID V21209 standard; DNA; 1664976 BP.
XX
AC V21209;
XX
DT 10-NOV-1998 (first entry)
XX
DE Methanococcus jannaschii circular chromosome.
XX
KW Methanococcus jannaschii; methanogenic archaeon; circular chromosome;
KW genome; autotrophic; extrachromosomal element; identification; ds.
XX
OS Methanococcus jannaschii.
XX
PN WO9807830-A2.
XX
PD 26-FEB-1998.
XX
PF 22-AUG-1997; 97WO-US14900.
XX
PR 22-AUG-1996; 96US-0024428.
XX
PA (GENO-) INST GENOMIC RES.
PA (UNII) UNIV ILLINOIS FOUND.
PA (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI Bult CJ, Smith HO, Venter JC, White OR, Woese CR;
XX
DR WPI; 1998-169145/15.
XX
PT Complete genome sequence of methano-genic archaeon, Methanoccoccus
PT jannaschii - useful in identification of M. jannaschii genome
PT fragment
XX
PS Claim 13; Page 152-585; 614pp; English.
XX
CC The present sequence represents the complete 1.66-megabase pair genome
CC sequence of the Methanococcus jannaschii circular chromosome. The
CC present invention describes M. jannaschii open reading frames from the
CC genome sequence. The invention also describes a computer based system
CC for identifying fragments of the M. jannaschii genome that are
CC homologous to target nucleotide sequences, comprising: (a) data storage
CC means comprising the nucleotide sequence of the 1664976, 58407 or 16550
CC bp sequence (see V21209, V21210 and V21211), or a nucleotide sequence at
CC least 99.9% identical to it; (b) search means for comparing a target
CC sequence to the nucleotide sequence of the data storage means to
CC identify a homologous sequence, and (c) retrieval means for obtaining
CC the homologous sequence. The method, which is based on whole genome
CC random sequencing of an autotrophic archaeon M. jannaschii, the genome
CC of which consists of 3 physically distinct elements, a large circular
CC chromosome (the 1664976 bp sequence given in V21209), a large circular
CC extra-chromosomal element (the 58407 bp sequence given in V21210), and a
CC small circular extra-chromosomal element (the 16550 bp sequence given in
CC V21211), can be used in the identification of M. jannaschii genome
CC fragment.
XX
SQ Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T; 101 other;

Query Match 53.7%; Score 20.4; DB 19; Length 1664976;
Best Local Similarity 71.1%; Pred. No. 1e+02;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy      1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38
          | |||| | ||||  | ||||||   || ||||||| |
Db 1455046 AAGCTTCTTCTCAGTGCTTAAACGAAGAACAAAAGCAT 1455009
```

RESULT 18
X06832
ID X06832 standard; cDNA; 1593 BP.
XX
AC X06832;
XX
DT 26-APR-1999 (first entry)
XX
DE Disease associated protein kinase DAPK-2 cDNA.
XX
KW DAPK-2; disease associated protein kinase; human; diagnosis;
KW therapy; adult respiratory distress syndrome; allergy; asthma;
KW arteriosclerosis; bronchitis; emphysema; hypereosinophilia;
KW myocardial inflammation; pericardial inflammation; anaemia;,
KW rheumatoid arthritis; Addison's disease; AIDS; atherosclerosis;
KW atopic dermatitis; dermatomyositis; diabetes mellitus;
KW glomerulonephritis; gout; Grave's disease; lupus erythematosus;
KW multiple sclerosis; myasthenia gravis; osteoarthritis;
KW osteoporosis; pancreatitis; polycystic kidney disease;
KW polymyositis; scleroderma; Sjorgren's syndrome;
KW autoimmune thyroiditis; cancer; infection; trauma;
KW cell proliferation; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 64..1410
FT /*tag= a
XX
PN WO9858052-A2.
XX
PD 23-DEC-1998.
XX
PF 19-JUN-1998; 98WO-US12813.
XX
PR 19-JUN-1997; 97US-0878989.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Bandman O, Corley NC, Goli SK, Guegler KJ, Hillman JL;
PI Lal P, Shah P;
XX
DR WPI; 1999-080952/07.
DR P-PSDB; W88433.
XX
PT New disease associated protein kinases - used to stimulate cell
PT proliferation and to treat the immune response and cancer
XX
PS Claim 5; Page 63; 93pp; English.
XX
CC This cDNA sequence codes for human disease associated protein
CC kinase DAPK-2 (see W88433). DAPK-2 cDNA was first identified in
CC the TBLYNOT01 cDNA library using a computer search for amino acid
CC alignments, and a consensus sequence was derived from the extended
CC and overlapping Incyte clones 40194/TBLYNOT01, 278198/TESTNOT03,
CC and 1683885/PROSNOT15. DAPK-2 shows 65% homology with the human
CC vaccinia virus related kinase VRK1 (GI 1827450), and is
CC associated with cDNA libraries which are immortalised or
CC cancerous and which show inflammatory or immune responses. The
CC invention provides disease associated protein kinases DAPK-1 to
CC DAPK-7 (see W88432-38) and cDNA clones encoding them (see
CC X06831-36 and X06882), as well as expression vectors, host cells,
CC agonists, antagonists and antibodies. The invention further
CC provides uses of such products in the diagnosis, prevention and
CC treatment of diseases associated with cell proliferation,
CC especially cancer or an immune response (claimed). Conditions
CC that may be treated include adult respiratory distress syndrome,
CC allergies, asthma, arteriosclerosis, bronchitis, emphysema,
CC hypereosinophilia, myocardial or pericardial inflammation,
CC rheumatoid arthritis, Addison's disease, AIDS, anaemia,

```
CC   atherosclerosis, various diseases of the digestive system, atopic
CC   dermatitis, dermatomyositis, diabetes mellitus, glomerulonephritis,
CC   gout, Grave's disease, lupus erythematosus, multiple sclerosis,
CC   myasthenia gravis, osteoarthritis, osteoporosis, pancreatitis,
CC   polycystic kidney disease, polymyositis, scleroderma, Sjorgren's
CC   syndrome, autoimmune thyroiditis, complications of cancer,
CC   extracorporeal circulation, viral, bacterial, fungal, parasitic,
CC   protozoal and helminthic infections, and trauma (disclosed).  The
CC   DAPK nucleic acids are also used in a method for detection of DAPK
CC   expression levels in a biological sample.
XX
SQ   Sequence 1593 BP; 568 A; 283 C; 309 G; 433 T; 0 other;


  Query Match             52.6%;  Score 20;  DB 20;  Length 1593;
  Best Local Similarity   72.2%;  Pred. No. 97;
  Matches   26;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       2 ggcttttgctcaaagtttaaactttgagcaaaagcc 37
           ||  | |||||| ||||||   | |||  |||||||
Db     759 ggtatgtgctcatagtttagcatatgatgaaaagcc 794


RESULT  19
X06832/c
ID   X06832 standard; cDNA; 1593 BP.
XX
AC   X06832;
XX
DT   26-APR-1999  (first entry)
XX
DE   Disease associated protein kinase DAPK-2 cDNA.
XX
KW   DAPK-2; disease associated protein kinase; human; diagnosis;
KW   therapy; adult respiratory distress syndrome; allergy; asthma;
KW   arteriosclerosis; bronchitis; emphysema; hypereosinophilia;
KW   myocardial inflammation; pericardial inflammation; anaemia;,
KW   rheumatoid arthritis; Addison's disease; AIDS; atherosclerosis;
KW   atopic dermatitis; dermatomyositis; diabetes mellitus;
KW   glomerulonephritis; gout; Grave's disease; lupus erythematosus;
KW   multiple sclerosis; myasthenia gravis; osteoarthritis;
KW   osteoporosis; pancreatitis; polycystic kidney disease;
KW   polymyositis; scleroderma; Sjorgren's syndrome;
KW   autoimmune thyroiditis; cancer; infection; trauma;
KW   cell proliferation; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             64..1410
FT                   /*tag=  a
XX
PN   WO9858052-A2.
XX
PD   23-DEC-1998.
XX
PF   19-JUN-1998;   98WO-US12813.
XX
PR   19-JUN-1997;   97US-0878989.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Bandman O,  Corley NC,  Goli SK,  Guegler KJ,  Hillman JL;
PI   Lal P,  Shah P;
XX
DR   WPI; 1999-080952/07.
DR   P-PSDB; W88433.
XX
PT   New disease associated protein kinases - used to stimulate cell
PT   proliferation and to treat the immune response and cancer
XX
PS   Claim 5; Page 63; 93pp; English.
XX
CC   This cDNA sequence codes for human disease associated protein
CC   kinase DAPK-2 (see W88433).  DAPK-2 cDNA was first identified in
CC   the TBLYNOT01 cDNA library using a computer search for amino acid
CC   alignments, and a consensus sequence was derived from the extended
CC   and overlapping Incyte clones 40194/TBLYNOT01, 278198/TESTNOT03,
CC   and 1683885/PROSNOT15.  DAPK-2 shows 65% homology with the human
CC   vaccinia virus related kinase VRK1 (GI 1827450), and is
CC   associated with cDNA libraries which are immortalised or
CC   cancerous and which show inflammatory or immune responses.  The
CC   invention provides disease associated protein kinases DAPK-1 to
CC   DAPK-7 (see W88432-38) and cDNA clones encoding them (see
CC   X06831-36 and X06882), as well as expression vectors, host cells,
CC   agonists, antagonists and antibodies.  The invention further
CC   provides uses of such products in the diagnosis, prevention and
CC   treatment of diseases associated with cell proliferation,
CC   especially cancer or an immune response (claimed).  Conditions
CC   that may be treated include adult respiratory distress syndrome,
CC   allergies, asthma, arteriosclerosis, bronchitis, emphysema,
CC   hypereosinophilia, myocardial or pericardial inflammation,
CC   rheumatoid arthritis, Addison's disease, AIDS, anaemia,
CC   atherosclerosis, various diseases of the digestive system, atopic
CC   dermatitis, dermatomyositis, diabetes mellitus, glomerulonephritis,
CC   gout, Grave's disease, lupus erythematosus, multiple sclerosis,
CC   myasthenia gravis, osteoarthritis, osteoporosis, pancreatitis,
CC   polycystic kidney disease, polymyositis, scleroderma, Sjorgren's
CC   syndrome, autoimmune thyroiditis, complications of cancer,
CC   extracorporeal circulation, viral, bacterial, fungal, parasitic,
CC   protozoal and helminthic infections, and trauma (disclosed).  The
CC   DAPK nucleic acids are also used in a method for detection of DAPK
CC   expression levels in a biological sample.
XX
SQ   Sequence 1593 BP; 568 A; 283 C; 309 G; 433 T; 0 other;


  Query Match             52.6%;  Score 20;  DB 20;  Length 1593;
  Best Local Similarity   72.2%;  Pred. No. 97;
  Matches   26;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       2 ggcttttgctcaaagtttaaactttgagcaaaagcc 37
           |||||||  ||| |   |||||| |||||| |  ||
Db     794 GGCTTTTCATCATATGCTAAACTATGAGCACATACC 759


RESULT  20
V46339
ID   V46339 standard; cDNA to mRNA; 1833 BP.
XX
AC   V46339;
XX
DT   11-NOV-1998  (first entry)
XX
DE   Serine threonine kinase VRK2 coding sequence.
XX
KW   Serine threonine kinase; VRK1; VRK2; B1R kinase; cell growth control;
KW   antitumour agent; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             76..1266
FT                   /*tag=  a
FT                   /product= VRK2
XX
PN   WO9829552-A1.
XX
PD   09-JUL-1998.
XX
PF   25-DEC-1997;   97WO-JP04855.
XX
PR   27-DEC-1996;   96JP-0357864.
XX
PA   (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
```

PI Nezu J, Oku A;
XX
DR WPI; 1998-388133/33.
DR P-PSDB; W64772.
XX
PT Serine-threonine kinase highly expressed in actively growing cells -
PT useful for development of cell growth inhibitors and antitumour
PT agents
XX
PS Claim 4; Page 26-30; 63pp; Japanese.
XX
CC This sequence encodes the serine threonine kinase VRK2 of the invention.
CC The protein is highly expressed in actively growing cells (such
CC as foetal organs), and has significant homology with the B1R kinase of
CC vaccinia virus. The kinase can be used as a substrate for the
CC screening of potential inhibitors, which can then be used in the control
CC of cell growth and as antitumour agents. Antisense DNA delivered via a
CC suitable vector can also be used for control of cell growth.
XX
SQ Sequence 1833 BP; 604 A; 351 C; 397 G; 481 T; 0 other;

Query Match 52.6%; Score 20; DB 19; Length 1833;
Best Local Similarity 72.2%; Pred. No. 98;
Matches 26; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy       2 ggcttttgctcaaagtttaaactttgagcaaaagcc 37
           ||  | |||||| ||||||   | |||  |||||||
Db    1006 ggtatgtgctcatagtttagcatatgatgaaaagcc 1041
```

Search completed: June 7, 2001, 18:10:34
Job time: 3492 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 7, 2001, 18:06:03 ; Search time 112.59 Seconds
(without alignments)
58.930 Million cell updates/sec

Title: US-09-601-561-3
Perfect score: 38
Sequence: 1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	20	52.6	1593	2	US-08-878-989-9	Sequence 9, Appli
c	2	20	52.6	1593	2	US-08-878-989-9	Sequence 9, Appli
	3	20	52.6	1593	4	US-09-272-796-9	Sequence 9, Appli
c	4	20	52.6	1593	4	US-09-272-796-9	Sequence 9, Appli
	5	19.6	51.6	6365	1	US-08-352-990-1	Sequence 1, Appli
c	6	19.6	51.6	6365	1	US-08-352-990-1	Sequence 1, Appli
	7	18.8	49.5	561	1	US-08-067-684-13	Sequence 13, Appl
c	8	18.8	49.5	561	1	US-08-067-684-13	Sequence 13, Appl
	9	18.8	49.5	561	1	US-08-008-898-13	Sequence 13, Appl
c	10	18.8	49.5	561	1	US-08-008-898-13	Sequence 13, Appl
	11	18.8	49.5	561	2	US-08-459-818-13	Sequence 13, Appl
c	12	18.8	49.5	561	2	US-08-459-818-13	Sequence 13, Appl
	13	18.8	49.5	561	2	US-08-889-666-13	Sequence 13, Appl
c	14	18.8	49.5	561	2	US-08-889-666-13	Sequence 13, Appl
	15	18.8	49.5	561	2	US-08-465-078-13	Sequence 13, Appl
c	16	18.8	49.5	561	2	US-08-465-078-13	Sequence 13, Appl
	17	18.8	49.5	561	2	US-08-725-776-13	Sequence 13, Appl
c	18	18.8	49.5	561	2	US-08-725-776-13	Sequence 13, Appl
	19	18.8	49.5	561	2	US-08-488-062-13	Sequence 13, Appl
c	20	18.8	49.5	561	2	US-08-488-062-13	Sequence 13, Appl
	21	18.8	49.5	561	3	US-08-228-208A-13	Sequence 13, Appl
c	22	18.8	49.5	561	3	US-08-228-208A-13	Sequence 13, Appl
	23	18.8	49.5	561	5	PCT-US95-06726-35	Sequence 35, Appl
c	24	18.8	49.5	561	5	PCT-US95-06726-35	Sequence 35, Appl
	25	18.8	49.5	3621	2	US-09-019-201A-1	Sequence 1, Appli
c	26	18.8	49.5	3621	2	US-09-019-201A-1	Sequence 1, Appli
	27	18.6	48.9	2254	1	US-08-153-848-27	Sequence 27, Appl
c	28	18.6	48.9	2254	1	US-08-153-848-27	Sequence 27, Appl
	29	18.6	48.9	2254	4	US-09-299-843A-27	Sequence 27, Appl
c	30	18.6	48.9	2254	4	US-09-299-843A-27	Sequence 27, Appl
	31	18.6	48.9	2254	5	PCT-US93-11153-27	Sequence 27, Appl
c	32	18.6	48.9	2254	5	PCT-US93-11153-27	Sequence 27, Appl
	33	18.6	48.9	3119	4	US-09-299-843A-31	Sequence 31, Appl
c	34	18.6	48.9	3119	4	US-09-299-843A-31	Sequence 31, Appl
	35	18.4	48.4	2757	2	US-08-627-254C-26	Sequence 26, Appl
c	36	18.4	48.4	2757	2	US-08-627-254C-26	Sequence 26, Appl
	37	18	47.4	2237	1	US-08-487-135B-1	Sequence 1, Appli
c	38	18	47.4	2237	1	US-08-487-135B-1	Sequence 1, Appli
	39	18	47.4	2237	2	US-08-915-972A-1	Sequence 1, Appli
c	40	18	47.4	2237	2	US-08-915-972A-1	Sequence 1, Appli
	41	18	47.4	2237	2	US-09-177-909-1	Sequence 1, Appli
c	42	18	47.4	2237	2	US-09-177-909-1	Sequence 1, Appli
	43	17.8	46.8	1848	1	US-08-075-193-3	Sequence 3, Appli
c	44	17.8	46.8	1848	1	US-08-075-193-3	Sequence 3, Appli
	45	17.8	46.8	1848	2	US-08-564-090A-3	Sequence 3, Appli
c	46	17.8	46.8	1848	2	US-08-564-090A-3	Sequence 3, Appli
	47	17.8	46.8	1848	5	PCT-US94-06698-3	Sequence 3, Appli
c	48	17.8	46.8	1848	5	PCT-US94-06698-3	Sequence 3, Appli
	49	17.8	46.8	2896	2	US-08-709-923-1	Sequence 1, Appli
c	50	17.8	46.8	2896	2	US-08-709-923-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT   1
US-08-878-989-9
; Sequence 9, Application US/08878989
; Patent No. 5885803
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Guegler, Karl G.
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Goli, Surya K.
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  DISEASE ASSOCIATED PROTEIN
;    TITLE OF INVENTION:  KINASES
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/878,989
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J J
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0321 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1593 base pairs
```

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  TBLYNOT01
;      CLONE:  40194
US-08-878-989-9


  Query Match             52.6%;  Score 20;  DB 2;  Length 1593;
  Best Local Similarity   72.2%;  Pred. No. 21;
  Matches   26;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 ggcttttgctcaaagtttaaactttgagcaaaagcc 37
          ||  | |||||| ||||||   | |||  |||||||
Db    759 GGTATGTGCTCATAGTTTAGCATATGATGAAAAGCC 794


RESULT   2
US-08-878-989-9/c
; Sequence 9, Application US/08878989
; Patent No. 5885803
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Guegler, Karl G.
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Goli, Surya K.
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  DISEASE ASSOCIATED PROTEIN
;    TITLE OF INVENTION:  KINASES
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/878,989
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J J
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0321 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1593 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  TBLYNOT01
;      CLONE:  40194
US-08-878-989-9


  Query Match             52.6%;  Score 20;  DB 2;  Length 1593;
  Best Local Similarity   72.2%;  Pred. No. 21;
  Matches   26;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 ggcttttgctcaaagtttaaactttgagcaaaagcc 37
          ||||||| ||| |   |||||| |||||| |  ||
Db    794 GGCTTTTCATCATATGCTAAACTATGAGCACATACC 759


RESULT   3
US-09-272-796-9
; Sequence 9, Application US/09272796
; Patent No. 6207148
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Guegler, Karl G.
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Goli, Surya K.
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  DISEASE ASSOCIATED PROTEIN
;    TITLE OF INVENTION:  KINASES
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/272,796
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/878,989
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J J
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0321 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1593 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  TBLYNOT01
;      CLONE:  40194
US-09-272-796-9


  Query Match             52.6%;  Score 20;  DB 4;  Length 1593;
  Best Local Similarity   72.2%;  Pred. No. 21;
  Matches   26;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 ggcttttgctcaaagtttaaactttgagcaaaagcc 37
          ||  | |||||| ||||||   | |||  |||||||
Db    759 GGTATGTGCTCATAGTTTAGCATATGATGAAAAGCC 794


RESULT   4
US-09-272-796-9/c
```

```
; Sequence 9, Application US/09272796
; Patent No. 6207148
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Guegler, Karl G.
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Goli, Surya K.
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  DISEASE ASSOCIATED PROTEIN
;    TITLE OF INVENTION:  KINASES
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/272,796
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/878,989
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J J
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0321 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1593 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  TBLYNOT01
;      CLONE:  40194
US-09-272-796-9


  Query Match             52.6%;  Score 20;  DB 4;  Length 1593;
  Best Local Similarity   72.2%;  Pred. No. 21;
  Matches   26;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      2 ggcttttgctcaaagtttaaactttgagcaaaagcc 37
          |||||||  ||| |   |||||| |||||| |  ||
Db    794 GGCTTTTCATCATATGCTAAACTATGAGCACATACC 759


RESULT   5
US-08-352-990-1
; Sequence 1, Application US/08352990
; Patent No. 5681562
;  GENERAL INFORMATION:
;    APPLICANT:  SOBOL, ROBERT E
;    APPLICANT:  GAGE, FRED H
;    APPLICANT:  ROYSTON, IVOR
;    APPLICANT:  FRIEDMAN, THEODORE
;    TITLE OF INVENTION:  LYMPHOKINE GENE THERAPY OF CANCER
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CAMPBELL AND FLORES
;      STREET:  4370 LA JOLLA VILLAGE DRIVE, SUITE 700
;      CITY:  SAN DIEGO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  91211
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/352,990
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/863,641
;      FILING DATE:  19920403
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CAMPBELL, CATHRYN
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-SD 9295
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-535-9001
;      TELEFAX:  619-535-8949
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6365 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  complement (1..6365)
;      OTHER INFORMATION:  /note= "Complementary strand of
;      OTHER INFORMATION:  pLXSN-RI-IL2"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2557..3351
US-08-352-990-1


  Query Match             51.6%;  Score 19.6;  DB 1;  Length 6365;
  Best Local Similarity   73.5%;  Pred. No. 38;
  Matches   25;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      5 ttttgctcaaagtttaaactttgagcaaaagcct 38
            ||| ||||||||   |||||||  |   || ||||
Db   1931 TTTAGCTCAAAGCAAAAACTTTCACTTAAGGCCT 1964


RESULT   6
US-08-352-990-1/c
; Sequence 1, Application US/08352990
; Patent No. 5681562
;  GENERAL INFORMATION:
;    APPLICANT:  SOBOL, ROBERT E
;    APPLICANT:  GAGE, FRED H
;    APPLICANT:  ROYSTON, IVOR
;    APPLICANT:  FRIEDMAN, THEODORE
;    TITLE OF INVENTION:  LYMPHOKINE GENE THERAPY OF CANCER
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CAMPBELL AND FLORES
;      STREET:  4370 LA JOLLA VILLAGE DRIVE, SUITE 700
;      CITY:  SAN DIEGO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  91211
;    COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/352,990
;       FILING DATE:
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/863,641
;       FILING DATE:  19920403
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  CAMPBELL, CATHRYN
;       REGISTRATION NUMBER:  31,815
;       REFERENCE/DOCKET NUMBER:  P-SD 9295
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  619-535-9001
;       TELEFAX:  619-535-8949
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6365 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  complement (1..6365)
;       OTHER INFORMATION:  /note= "Complementary strand of
;       OTHER INFORMATION:  pLXSN-RI-IL2"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  2557..3351
US-08-352-990-1


  Query Match             51.6%;  Score 19.6;  DB 1;  Length 6365;
  Best Local Similarity   73.5%;  Pred. No. 38;
  Matches   25;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaa 34
          |||| ||   | |||||||   ||||||| |||
Db   1964 AGGCCTTAAGTGAAAGTTTTTGCTTTGAGCTAAA 1931


RESULT   7
US-08-067-684-13
; Sequence 13, Application US/08067684
; Patent No. 5434131
;  GENERAL INFORMATION:
;    APPLICANT:  Linsley, Peter S.
;    APPLICANT:  Ledbetter, Jeffrey A.
;    APPLICANT:  Damle, Nitin K.
;    APPLICANT:  Brady, William
;    TITLE OF INVENTION:  CTLA4 RECEPTOR AND METHODS FOR ITS USE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheldon & Mak
;      STREET:  225 South Lake Avenue, Suite 900
;      CITY:  Pasadena
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  91101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/067,684
;      FILING DATE:  26-MAY-1993
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adriano, Sarah B.
;      REGISTRATION NUMBER:  34,470
;      REFERENCE/DOCKET NUMBER:  7848-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  310/312-9900
;      TELEFAX:  310/479-8340
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  Homo sapiens
US-08-067-684-13


  Query Match             49.5%;  Score 18.8;  DB 1;  Length 561;
  Best Local Similarity   68.4%;  Pred. No. 51;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38
            ||  || | |||| || |    ||||||||||||  ||
Db    424 AGCTTTCTCCTCACAGCTGTTTCTTTGAGCAAAATGCT 461


RESULT   8
US-08-067-684-13/c
; Sequence 13, Application US/08067684
; Patent No. 5434131
;  GENERAL INFORMATION:
;    APPLICANT:  Linsley, Peter S.
;    APPLICANT:  Ledbetter, Jeffrey A.
;    APPLICANT:  Damle, Nitin K.
;    APPLICANT:  Brady, William
;    TITLE OF INVENTION:  CTLA4 RECEPTOR AND METHODS FOR ITS USE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheldon & Mak
;      STREET:  225 South Lake Avenue, Suite 900
;      CITY:  Pasadena
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  91101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/067,684
;      FILING DATE:  26-MAY-1993
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adriano, Sarah B.
;      REGISTRATION NUMBER:  34,470
;      REFERENCE/DOCKET NUMBER:  7848-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  310/312-9900
;      TELEFAX:  310/479-8340
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
```

```
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  Homo sapiens
US-08-067-684-13


  Query Match             49.5%;  Score 18.8;  DB 1;  Length 561;
  Best Local Similarity   68.4%;  Pred. No. 51;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38
          ||  ||||||||||||    | || |||| | ||  ||
Db    461 AGCATTTTGCTCAAAGAAACAGCTGTGAGGAGAAAGCT 424


RESULT   9
US-08-008-898-13
; Sequence 13, Application US/08008898
; Patent No. 5770197
;  GENERAL INFORMATION:
;    APPLICANT:  Linsley, Peter S
;    APPLICANT:  Ledbetter, Jeffrey A
;    APPLICANT:  Damle, Nitin K
;    APPLICANT:  Brady, William
;    TITLE OF INVENTION:  CTLA4 RECEPTOR AND METHODS FOR ITS USE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheldon & Mak
;      STREET:  201 South Lake Avenue, Suite 800
;      CITY:  Pasadena
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  91101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/008,898
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/723,617
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandel, SaraLynn
;      REGISTRATION NUMBER:  31,853
;      REFERENCE/DOCKET NUMBER:  7848
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (818) 796-4000
;      TELEFAX:  (818) 795-6321
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..561
US-08-008-898-13


  Query Match             49.5%;  Score 18.8;  DB 1;  Length 561;
  Best Local Similarity   68.4%;  Pred. No. 51;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38
          ||  || | |||| || |    ||||||||||||  ||
Db    424 AGCTTTCTCCTCACAGCTGTTTCTTTGAGCAAAATGCT 461


RESULT  10
US-08-008-898-13/c
; Sequence 13, Application US/08008898
; Patent No. 5770197
;  GENERAL INFORMATION:
;    APPLICANT:  Linsley, Peter S
;    APPLICANT:  Ledbetter, Jeffrey A
;    APPLICANT:  Damle, Nitin K
;    APPLICANT:  Brady, William
;    TITLE OF INVENTION:  CTLA4 RECEPTOR AND METHODS FOR ITS USE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sheldon & Mak
;      STREET:  201 South Lake Avenue, Suite 800
;      CITY:  Pasadena
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  91101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/008,898
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/723,617
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandel, SaraLynn
;      REGISTRATION NUMBER:  31,853
;      REFERENCE/DOCKET NUMBER:  7848
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (818) 796-4000
;      TELEFAX:  (818) 795-6321
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..561
US-08-008-898-13


  Query Match             49.5%;  Score 18.8;  DB 1;  Length 561;
  Best Local Similarity   68.4%;  Pred. No. 51;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38
          ||  ||||||||||||    | || |||| | ||  ||
Db    461 AGCATTTTGCTCAAAGAAACAGCTGTGAGGAGAAAGCT 424


RESULT  11
US-08-459-818-13
; Sequence 13, Application US/08459818
; Patent No. 5851795
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Linsley, Peter S.
;    APPLICANT:  Ledbetter, Jeffrey A.
;    APPLICANT:  Damle, Nitin K.
;    APPLICANT:  Brady, William
;    TITLE OF INVENTION:  CTLA4 Receptor and Uses Thereof
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merchant & Gould
;      STREET:  11150 Santa Monica Blvd., Suite 400
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  FastSeq 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/459,818
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adriano, Sarah B.
;      REGISTRATION NUMBER:  34,470
;      REFERENCE/DOCKET NUMBER:  30436.35US02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  310-445-1140
;      TELEFAX:  310-445-9031
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..561
US-08-459-818-13


  Query Match             49.5%;  Score 18.8;  DB 2;  Length 561;
  Best Local Similarity   68.4%;  Pred. No. 51;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38
          ||  || | |||| || |    ||||||||||||  ||
Db    424 AGCTTTCTCCTCACAGCTGTTTCTTTGAGCAAAATGCT 461


RESULT  12
US-08-459-818-13/c
; Sequence 13, Application US/08459818
; Patent No. 5851795
;  GENERAL INFORMATION:
;    APPLICANT:  Linsley, Peter S.
;    APPLICANT:  Ledbetter, Jeffrey A.
;    APPLICANT:  Damle, Nitin K.
;    APPLICANT:  Brady, William
;    TITLE OF INVENTION:  CTLA4 Receptor and Uses Thereof
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merchant & Gould
;      STREET:  11150 Santa Monica Blvd., Suite 400
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  FastSeq 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/459,818
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adriano, Sarah B.
;      REGISTRATION NUMBER:  34,470
;      REFERENCE/DOCKET NUMBER:  30436.35US02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  310-445-1140
;      TELEFAX:  310-445-9031
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..561
US-08-459-818-13


  Query Match             49.5%;  Score 18.8;  DB 2;  Length 561;
  Best Local Similarity   68.4%;  Pred. No. 51;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38
          ||  |||||||||||||    | || |||| | ||  ||
Db    461 AGCATTTTGCTCAAAGAAACAGCTGTGAGGAGAAAGCT 424


RESULT  13
US-08-889-666-13
; Sequence 13, Application US/08889666
; Patent No. 5885579
;  GENERAL INFORMATION:
;    APPLICANT:  Linsley, Peter S.
;    APPLICANT:  Ledbetter, Jeffrey A.
;    APPLICANT:  Damle, Nitin K.
;    APPLICANT:  Brady, William
;    APPLICANT:  Kiener, Peter A.
;    TITLE OF INVENTION:  CTLA4 Receptor and Uses Thereof
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merchant & Gould
;      STREET:  11150 Santa Monica Blvd., Suite 400
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/889,666
;      FILING DATE:  08-JUL-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/375390
;      FILING DATE:  18-JAN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adriano, Sarah B.
;      REGISTRATION NUMBER:  34,470
;      REFERENCE/DOCKET NUMBER:  30436-35US01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  310-445-1140
;      TELEFAX:  310-445-9031
```

```
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..561
US-08-889-666-13


  Query Match             49.5%;  Score 18.8;  DB 2;  Length 561;
  Best Local Similarity   68.4%;  Pred. No. 51;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38
          ||  || | |||| || |    ||||||||||||  ||
Db    424 AGCTTTCTCCTCACAGCTGTTTCTTTGAGCAAAATGCT 461


RESULT  14
US-08-889-666-13/c
; Sequence 13, Application US/08889666
; Patent No. 5885579
;  GENERAL INFORMATION:
;    APPLICANT:  Linsley, Peter S.
;    APPLICANT:  Ledbetter, Jeffrey A.
;    APPLICANT:  Damle, Nitin K.
;    APPLICANT:  Brady, William
;    APPLICANT:  Kiener, Peter A.
;    TITLE OF INVENTION:  CTLA4 Receptor and Uses Thereof
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merchant & Gould
;      STREET:  11150 Santa Monica Blvd., Suite 400
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/889,666
;      FILING DATE:  08-JUL-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/375390
;      FILING DATE:  18-JAN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adriano, Sarah B.
;      REGISTRATION NUMBER:  34,470
;      REFERENCE/DOCKET NUMBER:  30436-35US01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  310-445-1140
;      TELEFAX:  310-445-9031
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..561
US-08-889-666-13


  Query Match             49.5%;  Score 18.8;  DB 2;  Length 561;
  Best Local Similarity   68.4%;  Pred. No. 51;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38
          ||  |||||||||||||    | || |||| | ||  ||
Db    461 AGCATTTTGCTCAAAGAAACAGCTGTGAGGAGAAAGCT 424


RESULT  15
US-08-465-078-13
; Sequence 13, Application US/08465078
; Patent No. 5885796
;  GENERAL INFORMATION:
;    APPLICANT:  Linsley, Peter S.
;    APPLICANT:  Ledbetter, Jeffrey A.
;    APPLICANT:  Damle, Nitin K.
;    APPLICANT:  Brady, William
;    APPLICANT:  Kiener, Peter A.
;    TITLE OF INVENTION:  CTLA4 Receptor and Uses Thereof
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merchant & Gould
;      STREET:  11150 Santa Monica Blvd., Suite 400
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/465,078
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/375390
;      FILING DATE:  18-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adriano, Sarah B.
;      REGISTRATION NUMBER:  34,470
;      REFERENCE/DOCKET NUMBER:  30436-35US01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  310-445-1140
;      TELEFAX:  310-445-9031
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..561
US-08-465-078-13


  Query Match             49.5%;  Score 18.8;  DB 2;  Length 561;
  Best Local Similarity   68.4%;  Pred. No. 51;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38
          ||  || | |||| || |    ||||||||||||  ||
Db    424 AGCTTTCTCCTCACAGCTGTTTCTTTGAGCAAAATGCT 461


RESULT  16
US-08-465-078-13/c
; Sequence 13, Application US/08465078
```

```
; Patent No. 5885796
;  GENERAL INFORMATION:
;    APPLICANT:  Linsley, Peter S.
;    APPLICANT:  Ledbetter, Jeffrey A.
;    APPLICANT:  Damle, Nitin K.
;    APPLICANT:  Brady, William
;    APPLICANT:  Kiener, Peter A.
;    TITLE OF INVENTION:  CTLA4 Receptor and Uses Thereof
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merchant & Gould
;      STREET:  11150 Santa Monica Blvd., Suite 400
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/465,078
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/375390
;      FILING DATE:  18-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adriano, Sarah B.
;      REGISTRATION NUMBER:  34,470
;      REFERENCE/DOCKET NUMBER:  30436-35US01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  310-445-1140
;      TELEFAX:  310-445-9031
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..561
US-08-465-078-13


  Query Match             49.5%;  Score 18.8;  DB 2;  Length 561;
  Best Local Similarity   68.4%;  Pred. No. 51;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38
          ||  |||||||||||||    | || |||| | ||  ||
Db    461 AGCATTTTGCTCAAAGAAACAGCTGTGAGGAGAAAGCT 424


RESULT  17
US-08-725-776-13
; Sequence 13, Application US/08725776
; Patent No. 5968510
;  GENERAL INFORMATION:
;    APPLICANT:  Linsley, Peter S.
;    APPLICANT:  Ledbetter, Jeffrey A.
;    APPLICANT:  Damle, Nitin K.
;    APPLICANT:  Brady, William
;    APPLICANT:  Kiener, Peter A.
;    TITLE OF INVENTION:  CTLA4 Receptor and Uses Thereof
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merchant & Gould
;      STREET:  11150 Santa Monica Blvd., Suite 400
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/725,776
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/375390
;      FILING DATE:  18-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adriano, Sarah B.
;      REGISTRATION NUMBER:  34,470
;      REFERENCE/DOCKET NUMBER:  30436-35US01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  310-445-1140
;      TELEFAX:  310-445-9031
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..561
US-08-725-776-13


  Query Match             49.5%;  Score 18.8;  DB 2;  Length 561;
  Best Local Similarity   68.4%;  Pred. No. 51;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38
          ||  || | |||| || |    ||||||||||||  ||
Db    424 AGCTTTCTCCTCACAGCTGTTTCTTTGAGCAAAATGCT 461


RESULT  18
US-08-725-776-13/c
; Sequence 13, Application US/08725776
; Patent No. 5968510
;  GENERAL INFORMATION:
;    APPLICANT:  Linsley, Peter S.
;    APPLICANT:  Ledbetter, Jeffrey A.
;    APPLICANT:  Damle, Nitin K.
;    APPLICANT:  Brady, William
;    APPLICANT:  Kiener, Peter A.
;    TITLE OF INVENTION:  CTLA4 Receptor and Uses Thereof
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merchant & Gould
;      STREET:  11150 Santa Monica Blvd., Suite 400
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/725,776
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US 08/375390
;      FILING DATE:  18-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adriano, Sarah B.
;      REGISTRATION NUMBER:  34,470
;      REFERENCE/DOCKET NUMBER:  30436-35US01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  310-445-1140
;      TELEFAX:  310-445-9031
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..561
US-08-725-776-13


  Query Match             49.5%;  Score 18.8;  DB 2;  Length 561;
  Best Local Similarity   68.4%;  Pred. No. 51;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38
            ||  ||||||||||||    | || |||| | ||  ||
Db    461 AGCATTTTGCTCAAAGAAACAGCTGTGAGGAGAAAGCT 424


RESULT  19
US-08-488-062-13
; Sequence 13, Application US/08488062
; Patent No. 5977318
;  GENERAL INFORMATION:
;    APPLICANT:  Linsley, Peter S.
;    APPLICANT:  Ledbetter, Jeffrey A.
;    APPLICANT:  Damle, Nitin K.
;    APPLICANT:  Brady, William
;    APPLICANT:  Kiener, Peter A.
;    TITLE OF INVENTION:  CTLA4 Receptor and Uses Thereof
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merchant & Gould
;      STREET:  11150 Santa Monica Blvd., Suite 400
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,062
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/375390
;      FILING DATE:  18-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adriano, Sarah B.
;      REGISTRATION NUMBER:  34,470
;      REFERENCE/DOCKET NUMBER:  30436-35US01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  310-445-1140
;      TELEFAX:  310-445-9031
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..561
US-08-488-062-13


  Query Match             49.5%;  Score 18.8;  DB 2;  Length 561;
  Best Local Similarity   68.4%;  Pred. No. 51;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38
            ||  ||||||||||||    | || |||| | ||  ||
Db    424 AGCTTTCTCCTCACAGCTGTTTCTTTGAGCAAAATGCT 461


RESULT  20
US-08-488-062-13/c
; Sequence 13, Application US/08488062
; Patent No. 5977318
;  GENERAL INFORMATION:
;    APPLICANT:  Linsley, Peter S.
;    APPLICANT:  Ledbetter, Jeffrey A.
;    APPLICANT:  Damle, Nitin K.
;    APPLICANT:  Brady, William
;    APPLICANT:  Kiener, Peter A.
;    TITLE OF INVENTION:  CTLA4 Receptor and Uses Thereof
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merchant & Gould
;      STREET:  11150 Santa Monica Blvd., Suite 400
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90025
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,062
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/375390
;      FILING DATE:  18-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Adriano, Sarah B.
;      REGISTRATION NUMBER:  34,470
;      REFERENCE/DOCKET NUMBER:  30436-35US01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  310-445-1140
;      TELEFAX:  310-445-9031
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  561 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..561
US-08-488-062-13


  Query Match             49.5%;  Score 18.8;  DB 2;  Length 561;
  Best Local Similarity   68.4%;  Pred. No. 51;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;
```

```
Qy       1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38
           ||  ||||||||||||    | || |||| | ||  ||
Db     461 AGCATTTTGCTCAAAGAAACAGCTGTGAGGAGAAAGCT 424
```

Search completed: June 7, 2001, 18:06:04
Job time: 4422 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 7, 2001, 17:41:30 ; Search time 1600.48 Seconds
(without alignments)
207.420 Million cell updates/sec

Title: US-09-601-561-3
Perfect score: 38
Sequence: 1 aggcttttgctcaaagtttaaactttgagcaaaagcct 38

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_gss_fun:*
177: em_gss_hum1:*
178: em_gss_hum2:*
179: em_gss_hum3:*
180: em_gss_hum4:*
181: em_gss_hum5:*
182: em_gss_hum6:*
183: em_gss_hum7:*
184: em_gss_hum8:*
185: em_gss_hum9:*
186: em_gss_inv1:*
187: em_gss_inv2:*
188: em_gss_inv3:*
189: em_gss_other:*
190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	22.6	59.5	457	206	AQ409763	AQ409763 HS_5095_B
c	2	22.6	59.5	457	206	AQ409763	AQ409763 HS_5095_B
	3	22.4	58.9	311	28	AV310837	AV310837 AV310837
c	4	22.4	58.9	311	28	AV310837	AV310837 AV310837
	5	22.4	58.9	402	228	B97021	B97021 F15L6TFB IG
c	6	22.4	58.9	402	228	B97021	B97021 F15L6TFB IG
	7	22.4	58.9	438	31	AV678544	AV678544 AV678544
c	8	22.4	58.9	438	31	AV678544	AV678544 AV678544
	9	22.4	58.9	760	217	AZ187266	AZ187266 SP_1008_B
c	10	22.4	58.9	760	217	AZ187266	AZ187266 SP_1008_B
	11	22.2	58.4	706	145	BF208904	BF208904 601871517
c	12	22.2	58.4	706	145	BF208904	BF208904 601871517
	13	22.2	58.4	791	166	BE379886	BE379886 601159565
c	14	22.2	58.4	791	166	BE379886	BE379886 601159565
	15	22.2	58.4	993	166	BE379806	BE379806 601159565
c	16	22.2	58.4	993	166	BE379806	BE379806 601159565
	17	21.8	57.4	423	207	AQ523622	AQ523622 HS_5209_A
c	18	21.8	57.4	423	207	AQ523622	AQ523622 HS_5209_A

	19	21.8	57.4	529	214	AQ966827	AQ966827 LERIM42TR
c	20	21.8	57.4	529	214	AQ966827	AQ966827 LERIM42TR
	21	21.8	57.4	558	31	AV677747	AV677747 AV677747
c	22	21.8	57.4	558	31	AV677747	AV677747 AV677747
	23	21.8	57.4	588	31	AV677314	AV677314 AV677314
c	24	21.8	57.4	588	31	AV677314	AV677314 AV677314
	25	21.8	57.4	771	229	CNS011PK	AL100562 Drosophil
c	26	21.8	57.4	771	229	CNS011PK	AL100562 Drosophil
	27	21.8	57.4	904	223	AZ536366	AZ536366 ENTCX45TF
c	28	21.8	57.4	904	223	AZ536366	AZ536366 ENTCX45TF
	29	21.6	56.8	451	173	BG096919	BG096919 EST461438
c	30	21.6	56.8	451	173	BG096919	BG096919 EST461438
	31	21.6	56.8	924	142	BE974415	BE974415 601680402
c	32	21.6	56.8	924	142	BE974415	BE974415 601680402
	33	21.2	55.8	295	128	BB244722	BB244722 BB244722
c	34	21.2	55.8	295	128	BB244722	BB244722 BB244722
	35	21.2	55.8	306	25	AV220207	AV220207 AV220207
c	36	21.2	55.8	306	25	AV220207	AV220207 AV220207
	37	21.2	55.8	447	161	BE007344	BE007344 PM2-BN014
c	38	21.2	55.8	447	161	BE007344	BE007344 PM2-BN014
	39	21.2	55.8	505	116	AW504296	AW504296 UI-HF-BN0
c	40	21.2	55.8	505	116	AW504296	AW504296 UI-HF-BN0
	41	21	55.3	275	165	BE318053	BE318053 NF062A04L
c	42	21	55.3	275	165	BE318053	BE318053 NF062A04L
	43	21	55.3	277	103	AI868005	AI868005 ta02g01.x
c	44	21	55.3	277	103	AI868005	AI868005 ta02g01.x
	45	21	55.3	298	164	BE198099	BE198099 ug79d12.x
c	46	21	55.3	298	164	BE198099	BE198099 ug79d12.x
	47	21	55.3	319	105	AL382080	AL382080 MtBC04F11
c	48	21	55.3	319	105	AL382080	AL382080 MtBC04F11
	49	21	55.3	402	220	AZ376506	AZ376506 1M0130G24
c	50	21	55.3	402	220	AZ376506	AZ376506 1M0130G24

ALIGNMENTS

```
RESULT   1
AQ409763
LOCUS       AQ409763      457 bp    DNA             GSS       17-MAR-1999
DEFINITION  HS_5095_B2_H06_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=671 Col=12 Row=P, DNA sequence.
ACCESSION   AQ409763
VERSION     AQ409763.1  GI:4431747
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 671  row: P  column: 12
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 457.
FEATURES             Location/Qualifiers
     source          1. .457
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=671 Col=12 Row=P"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      167 a     80 c     79 g    128 t      3 others
ORIGIN


  Query Match             59.5%;  Score 22.6;  DB 206;  Length 457;
  Best Local Similarity   75.7%;  Pred. No. 1.4e+02;
  Matches   28;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaagcc 37
          | ||||||| | ||   ||||| || ||||||| |||
Db    111 AAGCTTTTGGTGAAGAATTAAAATTAGAGCAAATGCC 147


RESULT   2
AQ409763/c
LOCUS       AQ409763      457 bp    DNA             GSS       17-MAR-1999
DEFINITION  HS_5095_B2_H06_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=671 Col=12 Row=P, DNA sequence.
ACCESSION   AQ409763
VERSION     AQ409763.1  GI:4431747
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 671  row: P  column: 12
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 457.
FEATURES             Location/Qualifiers
     source          1. .457
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=671 Col=12 Row=P"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
```

EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
BASE COUNT 167 a 80 c 79 g 128 t 3 others
ORIGIN

Query Match 59.5%; Score 22.6; DB 206; Length 457;
Best Local Similarity 75.7%; Pred. No. 1.4e+02;
Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 2 ggcttttgctcaaagtttaaactttgagcaaaagcct 38
||| ||||||| || ||||| || | ||||||| |
Db 147 GGCATTTGCTCTAATTTTAATTCTTCACCAAAAGCTT 111

RESULT 3
AV310837
LOCUS AV310837 311 bp mRNA EST 11-NOV-1999
DEFINITION AV310837 RIKEN full-length enriched, 8 days embryo Mus musculus cDNA clone 5730592L07 3', mRNA sequence.
ACCESSION AV310837
VERSION AV310837.1 GI:6363872
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 311)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al. 1999)
JOURNAL Unpublished (1999)
COMMENT Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center, Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL:http://genome.rtc.riken.go.jp/
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y., Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .311
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="5730592L07"
/clone_lib="RIKEN full-length enriched, 8 days embryo"
/sex="mixed"
/dev_stage="8 days embryo"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to Rot = 100.0 Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
BASE COUNT 68 a 66 c 61 g 115 t 1 others
ORIGIN

Query Match 58.9%; Score 22.4; DB 28; Length 311;
Best Local Similarity 81.2%; Pred. No. 1.6e+02;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 ggcttttgctcaaagtttaaactttgagcaaa 33
|||||||| | ||||||| ||||| |||||
Db 161 GGCTTTTGTTAAAAGTTTTTACTTTTTGCAAA 192

RESULT 4
AV310837/c
LOCUS AV310837 311 bp mRNA EST 11-NOV-1999
DEFINITION AV310837 RIKEN full-length enriched, 8 days embryo Mus musculus cDNA clone 5730592L07 3', mRNA sequence.
ACCESSION AV310837
VERSION AV310837.1 GI:6363872
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 311)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al. 1999)
JOURNAL Unpublished (1999)
COMMENT Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center, Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL:http://genome.rtc.riken.go.jp/
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y., Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,

```
           Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
           ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
           system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
           19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
           further details.
FEATURES             Location/Qualifiers
     source          1. .311
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="5730592L07"
                     /clone_lib="RIKEN full-length enriched, 8 days embryo"
                     /sex="mixed"
                     /dev_stage="8 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of subtraction to
                     Rot = 100.0 Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cloned into the XhoI and BamHI sites. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
BASE COUNT       68 a     66 c     61 g    115 t      1 others
ORIGIN


  Query Match             58.9%;  Score 22.4;  DB 28;  Length 311;
  Best Local Similarity   81.2%;  Pred. No. 1.6e+02;
  Matches   26;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      6 tttgctcaaagtttaaactttgagcaaaagcc 37
          |||||  |||||  ||||||| | ||||||||
Db    192 TTTGCAAAAAGTAAAAACTTTTAACAAAAGCC 161


RESULT   5
B97021
LOCUS       B97021        402 bp    DNA             GSS       31-MAR-1998
DEFINITION  F15L6TFB IGF Arabidopsis thaliana genomic clone F15L6, DNA
            sequence.
ACCESSION   B97021
VERSION     B97021.1  GI:2998267
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 402)
  AUTHORS   Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K.
            , Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter
            ,J.C.
  TITLE     A BAC End Sequence Database for Identifying Minimal Overlaps in
            Arabidopsis Genomic Sequencing. Update 3
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Steve Rounsley
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: rounsley@tigr.org
            Seq primer: M13-21
            Class: BAC ends
            High quality sequence stop: 402.
FEATURES             Location/Qualifiers
     source          1. .402
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="F15L6"
                     /clone_lib="IGF"
                     /sex="hermaphrodite"
                     /note="Vector: BeloBACII; Site_1: EcoRI; Site_2: EcoRI;
                     Produced by Thomas Altmann"
BASE COUNT       113 a     80 c     79 g    130 t
ORIGIN


  Query Match             58.9%;  Score 22.4;  DB 228;  Length 402;
  Best Local Similarity   81.2%;  Pred. No. 1.6e+02;
  Matches   26;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      3 gcttttgctcaaagtttaaactttgagcaaaa 34
            ||| ||||  | |||||||| ||||||||
Db    308 GCTGTTGCGGAGAGTTTAAAATTTGAGCAGAA 339


RESULT   6
B97021/c
LOCUS       B97021        402 bp    DNA             GSS       31-MAR-1998
DEFINITION  F15L6TFB IGF Arabidopsis thaliana genomic clone F15L6, DNA
            sequence.
ACCESSION   B97021
VERSION     B97021.1  GI:2998267
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 402)
  AUTHORS   Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K.
            , Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter
            ,J.C.
  TITLE     A BAC End Sequence Database for Identifying Minimal Overlaps in
            Arabidopsis Genomic Sequencing. Update 3
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Steve Rounsley
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: rounsley@tigr.org
            Seq primer: M13-21
            Class: BAC ends
            High quality sequence stop: 402.
FEATURES             Location/Qualifiers
     source          1. .402
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="F15L6"
                     /clone_lib="IGF"
                     /sex="hermaphrodite"
                     /note="Vector: BeloBACII; Site_1: EcoRI; Site_2: EcoRI;
                     Produced by Thomas Altmann"
BASE COUNT       113 a     80 c     79 g    130 t
ORIGIN
```

```
  Query Match            58.9%;  Score 22.4;  DB 228;  Length 402;
  Best Local Similarity  81.2%;  Pred. No. 1.6e+02;
  Matches   26;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      5 ttttgctcaaagtttaaactttgagcaaaagc 36
          || |||||||| |||||||| |  |||| |||
Db    339 TTCTGCTCAAATTTTAAACTCTCCGCAACAGC 308


RESULT   7
AV678544
LOCUS       AV678544      438 bp    mRNA            EST       05-OCT-2000
DEFINITION  AV678544 Nori Satoh unpublished cDNA library Ciona intestinalis
            cDNA clone rcitb1401 3', mRNA sequence.
ACCESSION   AV678544
VERSION     AV678544.1  GI:10116543
KEYWORDS    EST.
SOURCE      Ciona intestinalis.
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 438)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  TITLE     Expressed genes in Ciona intestinalis
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .438
                     /organism="Ciona intestinalis"
                     /db_xref="taxon:7719"
                     /clone="rcitb1401"
                     /clone_lib="Nori Satoh unpublished cDNA library"
                     /tissue_type="whole animal"
                     /dev_stage="tailbud"
BASE COUNT      170 a     68 c     77 g    123 t
ORIGIN


  Query Match            58.9%;  Score 22.4;  DB 31;  Length 438;
  Best Local Similarity  81.2%;  Pred. No. 1.7e+02;
  Matches   26;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      3 gcttttgctcaaagtttaaactttgagcaaaa 34
          ||||| | ||||| || |||| || |||||||
Db    197 GCTTTGGTTCAAATTTGAAACCTTAAGCAAAA 228


RESULT   8
AV678544/c
LOCUS       AV678544      438 bp    mRNA            EST       05-OCT-2000
DEFINITION  AV678544 Nori Satoh unpublished cDNA library Ciona intestinalis
            cDNA clone rcitb1401 3', mRNA sequence.
ACCESSION   AV678544
VERSION     AV678544.1  GI:10116543
KEYWORDS    EST.
SOURCE      Ciona intestinalis.
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 438)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  TITLE     Expressed genes in Ciona intestinalis
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .438
                     /organism="Ciona intestinalis"
                     /db_xref="taxon:7719"
                     /clone="rcitb1401"
                     /clone_lib="Nori Satoh unpublished cDNA library"
                     /tissue_type="whole animal"
                     /dev_stage="tailbud"
BASE COUNT      170 a     68 c     77 g    123 t
ORIGIN


  Query Match            58.9%;  Score 22.4;  DB 31;  Length 438;
  Best Local Similarity  81.2%;  Pred. No. 1.7e+02;
  Matches   26;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      5 ttttgctcaaagtttaaactttgagcaaaagc 36
          ||||||| || |||| || ||||| | |||||
Db    228 TTTTGCTTAAGGTTTCAAATTTGAACCAAAGC 197


RESULT   9
AZ187266
LOCUS       AZ187266      760 bp    DNA             GSS       30-AUG-2000
DEFINITION  SP_1008_B2_A07_T7A Strongylocentrotus purpuratus, purple sea urchin
            , sperm genomic BAC library Strongylocentrotus purpuratus genomic
            clone Plate=1008 Col=14 Row=B, DNA sequence.
ACCESSION   AZ187266
VERSION     AZ187266.1  GI:8370445
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 760)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
            ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and
            Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 1008  row: B  column: 14
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 760.
FEATURES             Location/Qualifiers
     source          1. .760
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="Plate=1008 Col=14 Row=B"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
BASE COUNT      224 a    136 c    153 g    247 t
ORIGIN


  Query Match            58.9%;  Score 22.4;  DB 217;  Length 760;
  Best Local Similarity  81.2%;  Pred. No. 1.8e+02;
```

```
  Matches   26;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      2 ggcttttgctcaaagtttaaactttgagcaaa 33
          ||| |||| | ||| |||||||||| || ||
Db    720 GGCGTTTGATTAAAATTTAAACTTTGGGCCAA 751


RESULT 10
AZ187266/c
LOCUS       AZ187266      760 bp    DNA             GSS       30-AUG-2000
DEFINITION  SP_1008_B2_A07_T7A Strongylocentrotus purpuratus, purple sea urchin
            , sperm genomic BAC library Strongylocentrotus purpuratus genomic
            clone Plate=1008 Col=14 Row=B, DNA sequence.
ACCESSION   AZ187266
VERSION     AZ187266.1  GI:8370445
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 760)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
            ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and
            Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 1008  row: B  column: 14
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 760.
FEATURES              Location/Qualifiers
     source           1. .760
                      /organism="Strongylocentrotus purpuratus"
                      /db_xref="taxon:7668"
                      /clone="Plate=1008 Col=14 Row=B"
                      /clone_lib="Strongylocentrotus purpuratus, purple sea
                      urchin, sperm genomic BAC library"
                      /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                      DH10B"
BASE COUNT      224 a    136 c    153 g    247 t
ORIGIN


  Query Match           58.9%;  Score 22.4;  DB 217;  Length 760;
  Best Local Similarity   81.2%;  Pred. No. 1.8e+02;
  Matches   26;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      6 tttgctcaaagtttaaactttgagcaaaagcc 37
          || || |||||||||||| ||| | |||| |||
Db    751 TTGGCCCAAAGTTTAAATTTTAATCAAACGCC 720


RESULT 11
BF208904
LOCUS       BF208904      706 bp    mRNA            EST       06-NOV-2000
DEFINITION  601871517F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:4092196 5',
            mRNA sequence.
ACCESSION   BF208904
VERSION     BF208904.1  GI:11102490
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 706)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM949  row: b  column: 05
            High quality sequence stop: 680.
FEATURES              Location/Qualifiers
     source           1. .706
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:4092196"
                      /clone_lib="NIH_MGC_53"
                      /tissue_type="carcinoma, cell line"
                      /lab_host="DH10B (T1 phage-resistant)"
                      /note="Organ: bladder; Vector: pDNR-LIB (Clontech);
                      Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                      ); Double-stranded cDNA was prepared from cell line RNA.
                      5' and 3' adaptors were used in cloning as follows: 5'
                      adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                      sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                      (where B = A, C, or G and N = A, C, G, or T).   Average
                      insert size 1.55 kb (range 0.9-4.0 kb).  15/15 colonies
                      contained inserts by PCR. This library was enriched for
                      full-length clones and was constructed by Clontech
                      Laboratories (Palo Alto, CA)."
BASE COUNT      220 a    123 c    131 g    232 t
ORIGIN


  Query Match           58.4%;  Score 22.2;  DB 145;  Length 706;
  Best Local Similarity   77.1%;  Pred. No. 2.2e+02;
  Matches   27;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaag 35
          ||  | || |||| | ||||| ||||||| |||||
Db    210 AGATTATTTCTCAGAATTTAATCTTTGAGGAAAAG 244


RESULT 12
BF208904/c
LOCUS       BF208904      706 bp    mRNA            EST       06-NOV-2000
DEFINITION  601871517F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:4092196 5',
            mRNA sequence.
ACCESSION   BF208904
VERSION     BF208904.1  GI:11102490
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 706)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
```

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM949 row: b column: 05
High quality sequence stop: 680.

```
FEATURES             Location/Qualifiers
     source          1. .706
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4092196"
                     /clone_lib="NIH_MGC_53"
                     /tissue_type="carcinoma, cell line"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: bladder; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); Double-stranded cDNA was prepared from cell line RNA.
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C, G, or T). Average
                     insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      220 a    123 c    131 g    232 t
ORIGIN


  Query Match             58.4%;  Score 22.2;  DB 145;  Length 706;
  Best Local Similarity   77.1%;  Pred. No. 2.2e+02;
  Matches   27;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       4 cttttgctcaaagtttaaactttgagcaaaagcct 38
           ||||| ||||||| ||||| | |||| || |  ||
Db     244 CTTTTCCTCAAAGATTAAATTCTGAGAAATAATCT 210


RESULT 13
BE379886
LOCUS       BE379886      791 bp    mRNA            EST       21-JUL-2000
DEFINITION  601159565F2 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3510832 5',
            mRNA sequence.
ACCESSION   BE379886
VERSION     BE379886.1  GI:9325251
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 791)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM196  row: b  column: 17
            High quality sequence stop: 674.
FEATURES             Location/Qualifiers
     source          1. .791
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3510832"
                     /clone_lib="NIH_MGC_53"
                     /tissue_type="carcinoma, cell line"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: bladder; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); Double-stranded cDNA was prepared from cell line RNA.
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C, G, or T). Average
                     insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      200 a    161 c    175 g    255 t
ORIGIN


  Query Match             58.4%;  Score 22.2;  DB 166;  Length 791;
  Best Local Similarity   77.1%;  Pred. No. 2.2e+02;
  Matches   27;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 aggcttttgctcaaagtttaaactttgagcaaaag 35
           ||| || |  ||| | ||||| ||||||| |||||
Db     463 AGGATTATTTTCAGAATTTAATCTTTGAGGAAAAG 497


RESULT 14
BE379886/c
LOCUS       BE379886      791 bp    mRNA            EST       21-JUL-2000
DEFINITION  601159565F2 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3510832 5',
            mRNA sequence.
ACCESSION   BE379886
VERSION     BE379886.1  GI:9325251
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 791)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM196  row: b  column: 17
            High quality sequence stop: 674.
FEATURES             Location/Qualifiers
     source          1. .791
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3510832"
                     /clone_lib="NIH_MGC_53"
                     /tissue_type="carcinoma, cell line"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: bladder; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); Double-stranded cDNA was prepared from cell line RNA.
                     5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C, G, or T). Average
                     insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      200 a    161 c    175 g    255 t
ORIGIN


  Query Match             58.4%;  Score 22.2;  DB 166;  Length 791;
  Best Local Similarity   77.1%;  Pred. No. 2.2e+02;
```

```
  Matches   27;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      4 cttttgctcaaagtttaaactttgagcaaaagcct 38
          ||||| ||||||| ||||| | |||  | || |||
Db    497 CTTTTCCTCAAAGATTAAATTCTGAAAATAATCCT 463


RESULT 15
BE379806
LOCUS       BE379806      993 bp    mRNA            EST       21-JUL-2000
DEFINITION  601159565T1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3510832 3',
            mRNA sequence.
ACCESSION   BE379806
VERSION     BE379806.1  GI:9325171
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 993)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM196  row: b  column: 17
            High quality sequence start: 29
            High quality sequence stop: 705.
FEATURES              Location/Qualifiers
     source           1. .993
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:3510832"
                      /clone_lib="NIH_MGC_53"
                      /tissue_type="carcinoma, cell line"
                      /lab_host="DH10B (T1 phage-resistant)"
                      /note="Organ: bladder; Vector: pDNR-LIB (Clontech);
                      Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                      ); Double-stranded cDNA was prepared from cell line RNA.
                      5' and 3' adaptors were used in cloning as follows: 5'
                      adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                      sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                      (where B = A, C, or G and N = A, C, G, or T).   Average
                      insert size 1.55 kb (range 0.9-4.0 kb).  15/15 colonies
                      contained inserts by PCR. This library was enriched for
                      full-length clones and was constructed by Clontech
                      Laboratories (Palo Alto, CA)."
BASE COUNT      314 a    188 c    187 g    303 t      1 others
ORIGIN


  Query Match             58.4%;  Score 22.2;  DB 166;  Length 993;
  Best Local Similarity   77.1%;  Pred. No. 2.3e+02;
  Matches   27;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      4 cttttgctcaaagtttaaactttgagcaaaagcct 38
          ||||| ||||||| ||||| | |||  | || |||
Db    584 CTTTTCCTCAAAGATTAAATTCTGAAAATAATCCT 618


RESULT 16
BE379806/c
LOCUS       BE379806      993 bp    mRNA            EST       21-JUL-2000
DEFINITION  601159565T1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3510832 3',
            mRNA sequence.
ACCESSION   BE379806
VERSION     BE379806.1  GI:9325171
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 993)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM196  row: b  column: 17
            High quality sequence start: 29
            High quality sequence stop: 705.
FEATURES              Location/Qualifiers
     source           1. .993
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:3510832"
                      /clone_lib="NIH_MGC_53"
                      /tissue_type="carcinoma, cell line"
                      /lab_host="DH10B (T1 phage-resistant)"
                      /note="Organ: bladder; Vector: pDNR-LIB (Clontech);
                      Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                      ); Double-stranded cDNA was prepared from cell line RNA.
                      5' and 3' adaptors were used in cloning as follows: 5'
                      adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                      sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                      (where B = A, C, or G and N = A, C, G, or T).   Average
                      insert size 1.55 kb (range 0.9-4.0 kb).  15/15 colonies
                      contained inserts by PCR. This library was enriched for
                      full-length clones and was constructed by Clontech
                      Laboratories (Palo Alto, CA)."
BASE COUNT      314 a    188 c    187 g    303 t      1 others
ORIGIN


  Query Match             58.4%;  Score 22.2;  DB 166;  Length 993;
  Best Local Similarity   77.1%;  Pred. No. 2.3e+02;
  Matches   27;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaaag 35
          ||| || |  ||| | ||||| ||||||| |||||
Db    618 AGGATTATTTTCAGAATTTAATCTTTGAGGAAAAG 584


RESULT 17
AQ523622
LOCUS       AQ523622      423 bp    DNA             GSS       11-MAY-1999
DEFINITION  HS_5209_A2_C07_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=785 Col=14 Row=E, DNA sequence.
ACCESSION   AQ523622
VERSION     AQ523622.1  GI:4771019
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
```

```
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 785  row: E  column: 14
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 423.
FEATURES              Location/Qualifiers
     source           1. .423
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=785 Col=14 Row=E"
                      /clone_lib="RPCI-11 Human Male BAC Library"
                      /sex="male"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      Male blood DNA was isolated from one randomly chosen donor
                      and partially digested with a combination of EcoRI and
                      EcoRI Methylase. Size selected DNA was cloned into the
                      pBACe3.6 vector at EcoRI sites"
BASE COUNT      126 a     71 c     99 g    127 t
ORIGIN


  Query Match             57.4%;  Score 21.8;  DB 207;  Length 423;
  Best Local Similarity   78.8%;  Pred. No. 2.8e+02;
  Matches   26;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 aggcttttgctcaaagtttaaactttgagcaaa 33
          |||||||| |||||| | ||  || |||| |||
Db    119 AGGCTTTTTCTCAAAATCTACTCTATGAGTAAA 151


RESULT  18
AQ523622/c
LOCUS       AQ523622      423 bp    DNA             GSS       11-MAY-1999
DEFINITION  HS_5209_A2_C07_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=785 Col=14 Row=E, DNA sequence.
ACCESSION   AQ523622
VERSION     AQ523622.1  GI:4771019
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 423)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 785  row: E  column: 14
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 423.
FEATURES              Location/Qualifiers
     source           1. .423
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=785 Col=14 Row=E"
                      /clone_lib="RPCI-11 Human Male BAC Library"
                      /sex="male"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      Male blood DNA was isolated from one randomly chosen donor
                      and partially digested with a combination of EcoRI and
                      EcoRI Methylase. Size selected DNA was cloned into the
                      pBACe3.6 vector at EcoRI sites"
BASE COUNT      126 a     71 c     99 g    127 t
ORIGIN


  Query Match             57.4%;  Score 21.8;  DB 207;  Length 423;
  Best Local Similarity   78.8%;  Pred. No. 2.8e+02;
  Matches   26;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      6 tttgctcaaagtttaaactttgagcaaaagcct 38
            ||| |||| ||  || | ||||| ||||||||
Db    151 TTTACTCATAGAGTAGATTTTGAGAAAAAGCCT 119


RESULT  19
AQ966827
LOCUS       AQ966827      529 bp    DNA             GSS       28-JAN-2000
DEFINITION  LERIM42TRB LERG Arabidopsis thaliana genomic clone LERIM42, DNA
            sequence.
ACCESSION   AQ966827
VERSION     AQ966827.1  GI:6794528
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 529)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: at@tigr.org
            For addtional information, see http://www.tigr.org/tdb/at/at.html
            Seq primer: TR
            Class: shotgun.
FEATURES              Location/Qualifiers
     source           1. .529
                      /organism="Arabidopsis thaliana"
                      /strain="Landsberg erecta"
                      /db_xref="taxon:3702"
                      /clone="LERIM42"
                      /clone_lib="LERG"
                      /note="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was
                      sheared to 0.4-0.7 Kbp before ligation."
BASE COUNT      154 a     93 c     88 g    194 t
ORIGIN
```

```
  Query Match             57.4%;  Score 21.8;  DB 214;  Length 529;
  Best Local Similarity   78.8%;  Pred. No. 2.9e+02;
  Matches   26;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      3 gcttttgctcaaagtttaaactttgagcaaaag 35
          |||||| |||||  || ||||||||  | ||||
Db     90 GCTTTTTCTCAAGTTTAAAACTTTGTTCTAAAG 122


RESULT  20
AQ966827/c
LOCUS       AQ966827      529 bp    DNA             GSS       28-JAN-2000
DEFINITION  LERIM42TRB LERG Arabidopsis thaliana genomic clone LERIM42, DNA
            sequence.
ACCESSION   AQ966827
VERSION     AQ966827.1  GI:6794528
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 529)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: at@tigr.org
            For addtional information, see http://www.tigr.org/tdb/at/at.html
            Seq primer: TR
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .529
                     /organism="Arabidopsis thaliana"
                     /strain="Landsberg erecta"
                     /db_xref="taxon:3702"
                     /clone="LERIM42"
                     /clone_lib="LERG"
                     /note="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was
                     sheared to 0.4-0.7 Kbp before ligation."
BASE COUNT      154 a     93 c     88 g    194 t
ORIGIN


  Query Match             57.4%;  Score 21.8;  DB 214;  Length 529;
  Best Local Similarity   78.8%;  Pred. No. 2.9e+02;
  Matches   26;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      4 cttttgctcaaagtttaaactttgagcaaaagc 36
          |||| |  |||||||| ||  ||||| ||||||
Db    122 CTTTAGAACAAAGTTTTAAACTTGAGAAAAAGC 90


Search completed: June  7, 2001, 17:41:32
Job time: 5370 sec
```